

**PSYCHOAKUSTISCHE UND NEUROPHYSIOLOGISCHE BEFUNDE
ZUR AUDITORISCHEN NACHVERDECKUNG
BEI DER WAHRNEHMUNG VON VERSCHLUSSLAUTEN**

Dissertation
zur Erlangung des Grades
Doktor der Philosophie (Dr. phil.)
der Universität Bremen

von
Helmut Spenner
aus Münster

~~1983~~

1984



Digitized by the Internet Archive
in 2026

**PSYCHOAKUSTISCHE UND NEUROPHYSIOLOGISCHE BEFUNDE
ZUR AUDITORISCHEN NACHVERDECKUNG
BEI DER WAHRNEHMUNG VON VERSCHLUSSLAUTEN**

Dissertation
zur Erlangung des Grades
Doktor der Philosophie (Dr. phil.)
der Universität Bremen

von
Helmut Spenner
aus Münster

1983

Die Arbeit hat dem Promotionsausschuß "Dr. phil" als Dissertation vorgelegen.

Gutachter: Prof. Dr. M. Stadler

Prof. Dr. O.D. Creutzfeldt

Tag der Prüfung: 6. 2. 1984

Inhaltsverzeichnis

	Seite
1. Einleitung	1
2. Neurophysiologische Befunde	13
2.1 Verdeckungsprozesse	14
2.2 Neurophysiologie und Sprache	18
3. Zu prüfende Annahmen	22
4. Stimuli und Methoden	24
4.1 Auswahl der Versuchswörter	24
4.2 Erstellung der Stimuli	25
4.2.1 Testwörter	25
4.2.2 Verschußlautreize	26
4.2.3 Variation der Maskiererpegel	27
4.2.4 Zusätzliche Stimuli	27
4.3 Psychoakustische Messungen	29
4.3.1 Reizdarbietung	29
4.3.2 Datenerhebung	30

Inhaltsverzeichnis (Forts.)

4.4	Neurophysiologische Messungen	30
4.4.1	Vorbereitung der Versuchstiere	30
4.4.2	Ableitungstechnik	32
4.4.3	Reizdarbietung	33
4.4.4	Ermittlung der Antwortraten	33
5.	Besprechung der Ergebnisse	35
5.1	Psychoakustische Messungen	35
5.1.1	Erste Versuchsreihe	35
5.1.2	Zweite Versuchsreihe	36
5.2	Neurophysiologische Messungen	38
5.2.1	Erste Versuchsreihe	38
5.2.2	Zweite Versuchsreihe	47
5.2.3	Antwortverhalten einzelner Zellen	54
6.	Diskussion	58
7.	Zusammenfassung	67
8.	Literaturverzeichnis	69

1. Einleitung

Derzeitige Theorien oder Modelle zur Wahrnehmung von Sprachlauten lassen sich grob in zwei Gruppen einteilen. Die der einen basieren auf der Annahme, daß eine auditorische, auf allgemeine Eigenschaften des Gehörs gegründete Analyse der Sprachsignale möglich ist, während die andere darüber hinaus sprachspezifische Wahrnehmungsprozesse auf phonetischer Ebene postuliert. (Übersicht : Paap, 1975).

Neben den phonetischen Systemen, die auf der Basis der Artikulation Kombinationen von Merkmalen zugrundelegen, die sich zwischen den Sprachlauten unterscheiden (vgl. Ladefoged, 1975), sind psychoakustische Grundlagen der Wahrnehmung von Sprachlauten nur spärlich bekannt. Im Mittelpunkt der Psychoakustik stehen vorwiegend Modelle zu Prozessen der auditorischen Peripherie. Die Untersuchungen auf diesem Gebiet beziehen sich in der Hauptsache auf Diskriminationsleistungen des Gehörs bzw. die Messung verschiedenartiger Schwellen. In der Regel werden relativ einfache und wohldefinierte Signale, normalerweise entlang nur einer Dimension, z.B. der Frequenz oder der Lautstärke, variiert, während Sprachlaute komplexer akustischer Struktur und in der Zeit stark veränderlich sind. Hingegen liegt das Interesse bei Arbeiten zur Sprachwahrnehmung darin, diejenigen Merkmale der Sprachlaute zu bestimmen, die ihre Identifikation ermöglichen, was zur Entwicklung von Methoden geführt hat, die von denen der klassischen Psychoakustik abweichen. (vgl. Pastore, 1981, Pisoni, 1979).

Die akustische Darstellung von Sprache als Zeitsignal und im Spektrogramm zeigt bedeutsame Amplitudenschwankungen und spektrale Variabilität. Ein Beispiel gibt Abb. 1, die



getrennt für 21 Frequenzkanäle die zeitlichen Schwankungen der akustischen Energie eines natürlich gesprochenen Satzes darstellt (n. Lewien, 1983). Die zeitlichen Abstände der Amplitudenmaxima im Sprachsignal entsprechen in etwa Silbenlänge (200 - 300 ms). Diese energiereichen Anteile sind den Vokalen als den Silbenkernen, relativ geringere Amplituden den Konsonanten zuzuordnen (vgl. Stevens, 1980).

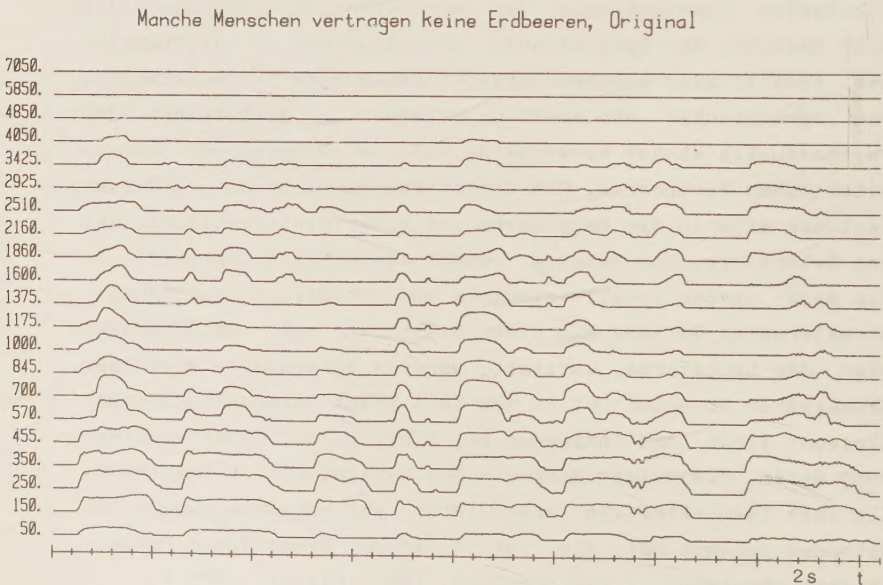


Abb. 1 : Darstellung des Amplitudenspektrums eines natürlich gesprochenen Satzes in Form von Einhüllenden für 21 Frequenzkanäle über die Zeit.

Obwohl Konsonanten (C) und Vokale (V) in einer Silbe als selbständige Sprachlaute wahrgenommen werden, sind ihre jeweiligen akustischen Entsprechungen im Zeitsignal nicht eindeutig trennbar, die Übergänge sind fließend.

Dies ist zurückzuführen auf kontinuierliche Veränderungen des Vokaltraktes bei der Erzeugung aufeinanderfolgender Sprachlaute (Koartikulation, Liberman et al., 1967) mit teilweise überlappenden, auch antizipatorischen, neuro-motorischen Befehlen an die einzelnen Artikulatoren (Lenneberg, 1967).

Im Gegensatz zu Verschußlauten sind Vokale und Frikative als Sprachlaute auch isoliert produzierbar und von genügender Länge, um im Sinne von Massaro (1972) als 'perceptual units' gelten zu können. Dabei können die frikativen Laute akustisch als modulierte Bandpaßbrauschen aufgefaßt werden.

Für Vokale lassen sich wegen der relativ langsamen spektralen Änderungen in einem mittleren Silbenbereich quasi-stationäre Anteile ausmachen. So finden sich in ihrer spektrographischen Darstellung mehrere energiereiche Bänder, die Formanten. Diese zeigen sich in Abb. 1 als Amplitudenmaxima in Bereichen verschiedener Frequenzkanäle. Das Band niedrigster Frequenz ist als Grundfrequenz (Frequenz der Anregung der Stimmlippen) im Wesentlichen eine Sprechereigenschaft und zwischen verschiedenen Vokalen wenig unterschieden.

Die beiden ersten Formanten hingegen spielen die bedeutende Rolle bei der Erkennung der Vokalqualität. So können aus entsprechenden Frequenzbändern synthetisierte Zwei-Formant-Annäherungen zuverlässig identifiziert werden.

Konfusionsstudien, Ähnlichkeitsschätzungen und mehrdimensionale Analysen der Unterschiede zwischen von verschiedenen Sprechern erzeugten Vokalen unterstreichen die essentielle

Bedeutung dieser Formanten bei der Vokalerkennung (Übersicht : Massaro, 1975).

Grundfrequenz und dritter Formant, ebenfalls über verschiedene Vokale recht konstant, scheinen zur Kompensation der zwischen Sprechern auftretenden Variabilität der akustischen Struktur von Vokalen beizutragen.

Solche Befunde weisen darauf hin, daß es nicht die absolute Lage der Formanten im Spektrum ist, die zur Identifikation verwertet wird. Welche komplizierteren Verarbeitungsweisen anzunehmen sind, ist noch umstritten (Karnickaya et al., 1975).

Die Produktion von Verschußlauten ist durch den vollständigen Verschuß des Vokaltraktes und seine anschließende plötzliche Öffnung charakterisiert. Diese Öffnung ist einem impulsartigen Vorgang vergleichbar. Dessen spektrale Lage ist vom Artikulationsort - dem Ort des Verschlusses im Vokaltrakt - abhängig. Aufgrund der Koartikulation von Konsonant und Vokal ergeben sich fließende Übergänge (Transienten) zwischen Öffnungsimpuls und benachbarten Vokalformanten. Für Verschußlaute am Silbenbeginn sind diese akustischen Ereignisse von sehr kurzer Dauer (10 - 40 ms, vgl. z.B. Stevens & Blumstein, 1978; Bailey & Summerfield, 1980).

Anders als Vokale sind Verschußlaute somit durch schnelle spektrale Änderungen gekennzeichnet.

Mit dem Ziel der Isolation der akustischen Merkmale von Verschußlauten, wurden in den Haskins-Labors mittels synthetischer Stimuli ausführliche Untersuchungen durchgeführt.

Es konnte gezeigt werden, daß Zwei-Formant-Annäherungen von Vokalen, die vor Erreichen des stationären Anteils derart frequenzmoduliert werden, daß sich Entsprechungen zu den For-

manttransienten natürlicher Sprache ergeben, als Konsonant-Vokal Silben gehört werden. Ob bilabial (/P/, /B/), alveolar (/T/, /D/) oder velar (/K/, /G/) artikulierte Verschlußlaute wahrgenommen werden, hängt dabei wesentlich vom Transienten des zweiten Formanten ab. Bei für verschiedene Vokalkontexte gegebener Formantstruktur erscheint dieser Transient je nach Verschlußlautgruppe auf einen bestimmten Ort des Frequenzspektrums gerichtet. Dieser 'locus' wurde als eventuell invariantes akustisches Merkmal für verschiedenerorts artikulierte Verschlußlaute dadurch ermittelt, daß transientenlose zweite Formanten in ihrer spektralen Lage variiert und diesen die jeweils gehörten Konsonanten zugeordnet wurden.

Mehrere Untersuchungen (vgl. Massaro, 1975) zeigen konstante 'loci' für /T/ und /D/ (um 1800 - 2000 Hz) unabhängig vom Vokalkontext. Für /P/ und /B/, besonders aber /K/ und /G/ sind die 'loci' je nach Vokalumgebung variabel, also keine invarianten akustischen Merkmale.

Zur Untersuchung der Bedeutung des bei der Öffnung des Vokaltraktes auftretenden akustischen Ereignisse wurden kurze schmalbandige Rauschimpulse verschiedener spektraler Lage, gefolgt von aus zwei stationären Formanten bestehenden synthetischen Vokalen dargeboten. Nur bei wenigen Rauschimpulsen mit den höchsten Mittenfrequenzen wird konsistent ein /T/ gehört, ansonsten entscheidet der Vokalkontext darüber, welcher der Verschlußlaute wahrgenommen wird (vgl. Massaro, 1975). So fanden Bailey & Summerfield (1980) eine Wechselwirkung zwischen der spektralen Lage des Impulses und der Frequenz des zweiten Formanten.

Zumindest mittels synthetischer Transienten und Impulse lassen sich also keine invarianten akustischen Merkmale für Verschlußlaute - mit Ausnahme der alveolaren - feststellen.

Für als gleich wahrgenommene phonetische Kategorien ergibt sich also eine kontextabhängige Variabilität der akustischen Merkmale, während der Artikulationsort recht invariant bleibt. Diese relative Stabilität von Produktionsparametern wird als Beleg für eine speziell phonetische Wahrnehmung von Sprachlauten, besonders Konsonanten, betrachtet, wobei artikulatorische Prozesse als Referenz angenommen werden (Motor-Theory of Speech Perception, Liberman et al., 1967, Liberman & Studdert-Kennedy, 1978).

An natürlich gesprochenen Silben durchgeführte Untersuchungen von Stevens & Blumstein (1978) ergeben eine alternative, auditorische Hypothese : Anstatt einzelne Formantenbergänge zu analysieren, verarbeitet das auditorische System die Muster oder Konfigurationen akustischer Ereignisse, die bei der Produktion von mit Verschlußlauten beginnenden Silben eintreten. Die Erkennung des jeweiligen Verschlußlautes geschieht anhand der für die verschiedenen Artikulationsorte charakteristischen groben Form des Spektrums der ersten 10 - 20 ms am Silbenbeginn. So wird den bilabialen Verschlußlauten ein entlang der Frequenzachse 'diffus fallendes', den alveolaren ein 'diffus steigendes', und den velaren ein 'kompaktes' Spektrum zugeschrieben. Mit entsprechenden 'akustischen Schablonen' konnten Blumstein & Stevens (1979) einen hohen Prozentsatz gesprochener Verschlußlaute korrekt klassifizieren, wenn auch Diehl (1981) vermutet, daß ihre Leistungsfähigkeit für zusammenhängende Sprachäußerungen nicht ausreicht. Experimentelle Daten zeigen, daß Verschlußlaute in variabler Vokalumgebung bei Darbietung derart kurzer Abschnitte des Zeitsignals (Tekieli & Cullinan, 1979) und anhand von Eigenschaften ihrer Kurzzeitspektra (Blumstein & Stevens, 1980) zuverlässig identifiziert werden.

Ein vergleichbarer Ansatz findet sich bei Klatt (1979), der die Möglichkeit direkten lexikalischen Zugangs von kombinierten Kurzzeitspektra aus diskutiert.

Trotz verschiedenartiger theoretischer Ansätze für die Wahrnehmung von Sprachlauten, speziell Konsonanten, ergeben alle diese Befunde, daß es die schnellen Veränderungen im Spektrum sind, die das für Verschußlaute relevante akustische Ereignis darstellen.

Ein Phänomen bei der Wahrnehmung von Verschußlauten, das im Sinne der Annahme eines phonetischen Wahrnehmungsmodus interpretiert wird, soll hier näher untersucht werden :

Dem für die Produktion dieser Gruppe von Konsonanten charakteristischen Verschuß des Vokaltraktes entspricht ein Intervall minimaler Energie (eine Pause) im akustischen Signal. Bailey & Summerfield (1980) maßen in /S/-Verschußlautsilben mit verschiedenen Vokalen für zwei (britische) Sprecher Pausenlängen zwischen 60 und 140 ms, die jeweils längsten für /P/, kürzere für /K/ und /T/.

Entfernt man die Pause aus dem Signal, wird trotz akustisch vollständiger Konsonant-Vokalsilbe der Verschußlaut nicht gehört, aus z.B. /STAHL/ wird /SCHAL/. Dieses Phänomen wurde für eine Vielzahl synthetischer und natürlicher Stimuli aufgezeigt, meist mit Frikativen, gefolgt von verschiedenen Verschußlauten und Vokalen, aber auch in anderen Phonemkontexten. Eine Reihe von Untersuchungen mit meist synthetischen Stimuli hat darüber hinaus zum Ergebnis, daß das Einfügen eines Intervalles zu Beginn eines Vokals zur Verschußlautwahrnehmung führt, auch wenn keine (weiteren) akustischen Merkmale wie z.B. Formanttransienten dafür vorhanden sind. Daher wird vermutet, daß die die Pause den

akustischen Hinweis auf den Verschluß des Vokaltraktes gibt, der wiederum das invariante (artikulatorische) Merkmal für das Vorhandensein eines Verschlußlautes im Sprachsignal darstellt (Liberman et al., 1967, Liberman & Pisoni, 1976, Marcus, 1978, Dorman et al., 1979). Spenner & Schreiner (1980) konnten allerdings keine Verschlußlautwahrnehmung durch Einfügen eines Intervalles bei natürlichen Äußerungen des englischen /SIN/ herbeiführen, auch ergeben weitere Versuche (unveröffentlicht) ein uneinheitliches Bild.

Die Stimuli, bei denen ein akustisch im Signal vorhandener Verschlußlaut bei ungenügender Länge des vorausgehenden Intervalls nicht gehört wird, entsprechen in ihrer Struktur der Anordnung von Signalen in psychoakustischen Nachverdeckungsexperimenten, in denen der Einfluß eines Signals auf die Wahrnehmbarkeit eines in kurzem Abstand folgenden zweiten Signals untersucht wird.

Psychoakustische Verdeckungsprozesse zeigen sich als teilweise oder vollständig verminderte Hörbarkeit eines (Test-) Schalles, verursacht durch einen Störschall (Maskierer). Je nach zeitlicher Lage des Testschalles zum Maskierer spricht man von Vor-, Simultan- oder Nachverdeckung.

Zur quantitativen Beschreibung dieser Vorgänge wird in der Regel die Mithörschwelle von reinen Tönen (manchmal auch von Schmalbandrauschen) ermittelt, also jener Pegel des Testtones, der gerade hoch genug ist, um diesen in Anwesenheit des Maskiersignales (Breit- und Schmalbandrauschen, Töne) wahrnehmen zu können (Zwicker, 1982).

Die Effektivität des Maskierers ist dabei von seiner Amplitude, dem Ausmaß spektraler Gemeinsamkeiten beider Signale, ihrer relativen Dauer und im Falle nichtsimultaner Verdeckung

ihrem zeitlichen Abstand abhängig.

Bei breitbandigen Störsignalen (weißes Rauschen, gleichmäßig verdeckendes Rauschen) kommt es zur Verdeckung über den gesamten Hörbereich. An dessen unteren und oberen Grenzen geht die Mithörschwelle in die Ruhehörschwelle über. Ein Anstieg des Maskiererpegels bewirkt eine gleichwertige Anhebung der Mithörschwelle (mit Einschränkungen beim weißen Rauschen).

Schmalbandige und tonale Verdeckungsreize sind am wirksamsten für Testsignale gleicher (Mitten-) Frequenz, zu niederen und höheren Frequenzen wird das Ausmaß der Verdeckung geringer. Die Verläufe der Mithörschwellen für Testsignale verschiedener Frequenz sind asymmetrisch, ihr Abfall ist zu hohen Frequenzen hin langsamer als zu niedrigen. Diese Asymmetrie wird mit wachsenden Maskiererpegeln ausgeprägter, der Frequenzbereich, innerhalb dessen Testsignale verdeckt werden, breiter. Mithörschwellen von über 200 ms langen Testsignalen sind von deren Dauer unabhängig, mit kürzerer steigen sie an.

Diese Befunde zur Simultanverdeckung (Zwicker & Feldtkeller, 1967) gelten allgemein auch für die nichtsimultanen Verdeckungsvorgänge. Zusätzlicher Parameter ist die Verzögerungszeit, also der Abstand des Testsignals vom Störschall.

Bei der Nachverdeckung entspricht die Nachhörschwelle bei kurzen Verzögerungszeiten der Mithörschwelle bei Simultanverdeckung, für Verzögerungszeiten ab 200 ms - maskiererpegelunabhängig - der Ruhehörschwelle.

Bei kurzen Maskierern wird die Ruhehörschwelle schon für kürzere Verzögerungszeiten erreicht, so bei einem Maskierer von 5 ms Dauer bei bereits 50 ms (Zwicker, 1982).

Im Falle der Vorverdeckung liegt das Testsignal vor dem Störschall, die Verzögerungszeit wird negativ. Im Vergleich

zur Nachverdeckung ist ihre Wirkung zeitlich wenig ausgedehnt. Man findet Verzögerungszeiten von -10 bis -20 ms bis zum Erreichen der Ruhehörschwelle.

Parallelen bei Sprachreizen finden sich in den Zeit- und Amplitudenverhältnissen der sie konstituierenden Laute.

Energieschwache Anteile des Verschlußlaut-Vokalüberganges sind von relativ energiereichen und ausgedehnten Lauten umgeben, die Größenordnung von Intervallängen und Verschlußlautdauer liegt im Bereich der Verzögerungszeiten, bei denen Nach- und Vorverdeckung vorkommen.

Gegenüber den in den klassischen Verdeckungsexperimenten benutzten Schallen allerdings ergeben sich für die Sprachlaute Unterschiede. Sind Frikative als Nachverdeckungsreize einem Bandpaßrauschen noch vergleichbar, so sind die CV-übergänge hinsichtlich ihrer spektralen Komplexität von den als Testreizen üblichen Ton- oder Rauschimpulsen bedeutsam unterschieden, die Verschlußlaute als solche nicht isolierbar.

Bietet man statt vollständiger Silben kurze Abschnitte der Verschlußlautübergänge als Testreize mit verschiedenen Verzögerungszeiten dar, so stellt man fest, daß sie in jedem Fall als Geräusch hörbar sind. Dorman et al. (1979), die entsprechende synthetische Formanttransienten von 50 ms Länge zur Prüfung der Nachverdeckungshypothese verwandten, sehen diese dadurch widerlegt. Nach ihrer Ansicht werden sie als nichtsprachliche Reize empfunden und auditorisch, dieselben Signale als Teile ganzer Silben phonetisch verarbeitet. In letzterem Fall widerspreche die Information eines fehlenden Intervalles der Artikulation eines Verschlußlautes, der trotz akustisch ausreichender Intensität demnach nicht gehört wird.

Dem ist entgegenzuhalten, daß diese Silbenanfänge ihren

größeren Energieanteil im Bereich der Transienten der Vokalformanten haben und daß ihre zeitliche Ausdehnung zu groß ist, um der Forderung zu genügen, daß das zu verdeckende Signal im Vergleich zu den vorkommenden Verzögerungszeiten kurz sein sollte (Zwicker & Feldtkeller, 1967). Es ist außerdem denkbar, daß für Verschußlaute unerläßliche Anteile der Nachverdeckung unterliegen, und andere nicht relevante Komponenten des Überganges, die bei Silbenabschnitten gehört werden, bei vollständigen Silben durch spektral verwandete des folgenden Vokals vorverdeckt werden.

Diese Komplexität der Sprachlaute verhindert den direkten Nachweis von Verdeckungsprozessen über die Messung von Mithörschwellen.

Ein naheliegender Weg, sprachspezifische und auditorische Vorgänge experimentell zu trennen, liegt in der Anwendung geeigneter Sprachlaute im Tierversuch.

Wenn hier dem Humanexperiment entsprechende Ergebnisse gefunden werden, sollten sie durch Eigenschaften des Gehörs erklärbar sein, weil eine mögliche Referenz zu Vorgängen auf der Ebene der Sprachproduktion beim Tier nicht gegeben ist.

Hier sind mit einigem Erfolg als sprachspezifisch aufgefaßte Leistungen menschlicher Wahrnehmung auf Verhaltensebene repliziert worden.

Diese betreffen die sogenannte 'kategorische Wahrnehmung' (Studdert-Kennedy 'et al., 1970) von Sprachlauten :

Bietet man eine geeignete Reihe physikalisch gleichabständiger synthetischer Stimuli an, so erhält man im Identifikationsversuch einen Phonemübergang von z.B. /BA/ nach /DA/. Sollen benachbarte Stimuli dieser Reihe diskriminiert werden, stellt sich heraus, daß die Unterscheidungsleistung in der

Nähe des vorher gemessenen Übergangs, also zwischen phonetischen Kategorien, weit besser ist als bei weiter abgelegenen Stimuluspaaren, die innerhalb der phonetischen Kategorien liegen, obwohl die Paare jeweils gleiche 'akustische Distanz' aufweisen.

Eine Reihe von Tierversuchen mit verschiedenen Species, z.B. Chinchillas und Affen, ergibt bei verschiedenen Verhaltenskriterien Resultate, die dieser kategorischen Wahrnehmung entsprechen. (vgl. Kuhl, 1979, Pisoni, 1979, Snowden, 1979, Kuhl, 1981).

Obwohl die Sprachspezifität kategorischer Wahrnehmung inzwischen auch deshalb zweifelhaft ist, weil sie ebenfalls bei nichtsprachlichen Signalen gefunden wird (z.B. Cutting & Rosner, 1974), ist deutlich, daß Tierversuche besonders bei psychoakustisch schwer zugänglichen Phänomenen einen sinnvollen Ansatz darstellen, um kritische Untersuchungen durchzuführen, die zwischen auditorischen und produktionsabhängigen Vorgängen bei der Wahrnehmung von Sprachlauten unterscheiden.

Das diesem Ansatz verwandte Ziel der vorliegenden Untersuchung ist es, anhand neurophysiologischer Messungen in der Hörbahn der Katze zu zeigen, daß die Antworten einzelner Nervenzellen auch bei komplexen Signalen entsprechend einer auditorischen Verdeckungsannahme ausfallen, das Phänomen der 'Verschlußlautunterdrückung' also als in Eigenschaften des Gehörsinnes begründet zu deuten ist.

2. Neurophysiologische Befunde

Eine Voraussetzung dafür ist, daß in der Hörbahn auftretende Prozesse den psychoakustischen Vorgängen vergleichbar sind. Es ist also vorab zu zeigen, daß auf der Ebene von Einzelzellen Verdeckungsvorgänge auftreten, und wie diese sich zu den psychophysischen Gegebenheiten verhalten. Des weiteren ist das Verhalten von Nervenzellen bei der Darbietung von Sprachlauten von Bedeutung : Insbesondere ist eine gute Darstellung der als Testreize vorgesehenen Verschlußlaute notwendig.

Als Maß für die Reaktivität der Nervenzellen auf bestimmte Reize werden in den meisten Untersuchungen Antwortraten angegeben. Zu ihrer Ermittlung werden die Reize wiederholt präsentiert. Die Alles - oder - Nichts - Antworten (Spikes) der sensiblen Neurone werden zeitsynchron über die Reizdarbietungen aufaddiert und diese Summen als sogenanntes Peri-Stimulus-Time Histogramm dargestellt.

Die den Stimuli zugeordneten Spikesummen ergeben relativ zur Zeiteinheit die jeweilige Antwortrate (vgl. Creutzfeldt & Probst, 1974). Das Verhalten solcher Raten in neurophysiologischen Verdeckungsexperimenten und bei der Darbietung von Sprachlauten ist Gegenstand der folgenden Darstellung neurophysiologischer Ergebnisse, um Hinweise zur Übertragung der verschiedenen Reiztypen und auf einen geeigneten Ableitungs-ort in der Hörbahn, zu erhalten.

2.1 Verdeckungsprozesse

Harris & Dallos (1979) führten eine Untersuchung zur Nachverdeckung an einzelnen Fasern des Hornervens von Chinchillas durch.

Die Testreize sind jeweils Töne der Bestfrequenz der Faser, also der Frequenz, bei der die geringste Intensität notwendig ist, um eine überschwellige Antwort zu erhalten.

Nachverdeckung zeigt sich hier - bei überschwelligen Testtönen - als eine Verminderung der Rate, mit der die Faser auf den Testreiz antwortet. Diese Verminderung ist umso ausgeprägter, je stärker die Antwort auf den Verdeckungsreiz ausfällt, wobei bis zu Verzögerungszeiten von etwa 50 ms die Verhältnisse durch einen linearen Zusammenhang gut darstellbar sind.

Betrag und Zeitkonstante der Nachverdeckung sind nach diesen Befunden nicht vom physikalischen Pegel des Maskierers direkt abhängig, sondern von dem Niveau der Erregung, das er auslöst. Für hohe Maskiererpegel kommt es zur Sättigung der Faserantwort schon bei Intensitäten weit unterhalb der Grenze des Hörbereichs. Dieser relativ begrenzte Dynamikbereich der auditorischen Peripherie wurde für eine Vielzahl unterschiedlicher Stimuli festgestellt (vgl. Evans, 1981 a).

Das Ausmaß an Verdeckung verhält sich parallel : zum Sättigungsbereich hin haben Änderungen des Maskiererpegels abnehmenden Einfluß auf den Betrag der Nachverdeckung.

Die kritische Dauer des Verdeckungsreizes liegt bei ca. 100 ms. Ihre weitere Zunahme hat keine steigernde Wirkung auf das Ausmaß der Antwortreduktion.

Die Erholungskurven der Antwort auf den Testreiz sind nach Harris & Dallos recht gut durch exponentielle Funktionen anzunähern. Ihre Verläufe und Zeitkonstanten sind von der Best-

frequenz der jeweiligen Faser und der spektralen Zusammensetzung des Maskierers unabhängig. Sie werden aber entscheidend von dessen Pegel beeinflusst. Wächst der Maskiererpegel, so werden auch die Zeitkonstanten der Erholungskurven größer, die Verzögerungszeiten, für die keine Verdeckung mehr festgestellt wird, sind verlängert.

Aufgrund dieser Befunde, insbesondere wegen des wesentlichen Einflusses der Größe der zeitlich früheren Faserantwort, wird Nachverdeckung in diesem peripheren Bereich der Hörbahn als Adaptationsvorgang verstanden (Harris & Dallos, 1979, Smith, 1977).

Betreffend den Einfluß des Maskiererpegels bei kurzen Reizabständen besteht eine deutliche qualitative Ähnlichkeit mit psychoakustischen Verdeckungsvorgängen, quantitative Übereinstimmungen finden sich für den Einfluß der Dauer des Maskierreizes, der im psychoakustischen Experiment bei vergleichbarer Ausdehnung maximal wirksam wird.

Unterschiede betreffen die Zeitkonstanten der Schwellen- bzw. Erholungskurven. So durchläuft bei Erhöhung des Maskiererpegels die Mithörschwellenkurve einen größeren Intensitätsbereich bis zum Erreichen der Ruhehörschwelle, während die dazu notwendige Verzögerungszeit gleich bleibt (Zwicker & Feldtkeller, 1967). Diese Verzögerungszeit nimmt bei den neurophysiologischen Erholungskurven zu, gleichermaßen ihre Zeitkonstanten. Dieser Unterschied scheint im Corpus geniculatum mediale (CGM), einer letzten Schaltstellen vor der Hörrinde, etwas geringer auszufallen (Schreiner, 1980 a).

In Anlehnung an ein weiteres psychoakustisches Verdeckungsphanomen untersuchte Schreiner (1980 b) die Verarbeitung alternierender Signale im CGM des Meerschweinchens.

Er bediente sich des Paradigmas der 'Pulsationsschwelle' : Bei angemessenen Frequenz- und Amplitudenverhältnissen wird bei alternierend angebotenen Maskier- und Testreizen der Testton als kontinuierlich und nicht - wie physikalisch gegeben - als pulsierend wahrgenommen. Der Übergang zwischen diesen Eindrücken ist die Pulsationsschwelle. Es kommt also zur Verdeckung zeitlicher Merkmale eines Signals aufgrund der Anwesenheit eines anderen.

Für (weiße) Rauschpulse als Maskierer und Testtöne bei der Bestfrequenz des jeweiligen Neurons findet Schreiner als neurophysiologische Entsprechung eine Unterdrückung der phasischen Zellantwort auf das Einsetzen der Tonpulse, also eine Abnahme der Markierung zeitlicher Signaleigenschaften bei wachsendem Maskiererpegel. Die quantitativen Einflüsse verschiedener Störschallpegel sind für psychoakustische und neurophysiologische Messungen vergleichbar.

Ähnlich gute Entsprechungen finden sich für die Pegelabhängigkeit von Mithörschwellen und Zellantwortschwellen bei nichtsimultaner Verdeckung (Abb. 2, n. Schreiner, 1980 b).

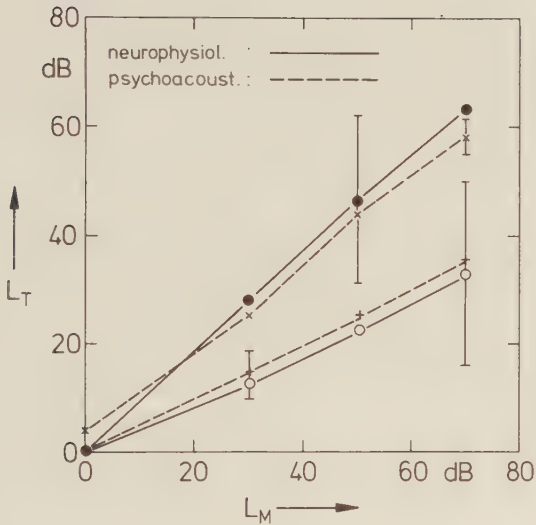


Abb. 2 : Die beiden oberen Kurven zeigen Pulsationsschwellen im Vergleich zu Schwellen phasischer Zellantworten, die unteren Mithör- und Zellantwortschwellen. L_M = Maskierpegel, L_T : Testtonpegel

Neurophysiologische Verdeckung zeigt sich also als Verschiebung von Schwellen oder als Verminderung von Antwortraten, was auch Untersuchungen mit überschwelligen Testreizen möglich macht. Die Übereinstimmung der Vorgänge mit psychoakustischer Verdeckung erscheint dabei auf höheren Ebenen der Hörbahn besser zu sein als in der auditorischen Peripherie.

2.2 Neurophysiologie und Sprache

Zur neurophysiologischen Darstellung von Sprachlauten liegen nur relativ wenige experimentelle Arbeiten vor.

Die meisten Arbeiten betreffen die auditorische Peripherie, vornehmlich den Hornnerven. Hier wurde insbesondere die Verarbeitung natürlicher und synthetischer Vokale untersucht, vereinzelt auch die von Silben.

Zusätzlich ergeben sich Hinweise anhand der Übertragung von in verschiedenem Maße sprachverwandten komplexen Signalen auf allen, vorwiegend aber übergeordneten Ebenen der Hörbahn. Auch die Verarbeitung artspezifischer Vokalisationen durch cortikale und subcortikale Strukturen liefert einige Erkenntnisse. Im Zusammenhang mit Verschlusslauten ist besonders die Darstellung spektraler Veränderungen, z.B. frequenz- und amplitudenmodulierter Signale zu berücksichtigen.

Mit einer Varietät sprachähnlicher Laute untersuchte Delgutte (1980) an betäubten Katzen die Entladungsmuster des Hornnerven.

Als Entsprechung zu den schnellen Amplitudenänderungen, die beim Wechsel von Konsonanten und Vokalen auftreten, wählte er Rauschimpulse mit verschiedenen langen Anstiegszeiten, analog zur Formantstruktur von Vokalen eine synthetische Annäherung aus Grundfrequenz und erstem Formanten. In beiden Fällen trat schon bei sprachüblichen (relativ geringen) Intensitäten eine Sättigung der Antworten ein. Amplitudensteigerungen haben also keine Zunahme der Aktivität des Nerven in Form von Entladungsraten mehr zur Folge, was nahelegt, daß die Antwortraten auf diesem Niveau der Hörbahn zur Codierung von Sprachlauten wenig beitragen.

Diese Ansicht unterstützen die Messungen von Young & Sachs (1979), die Antwortraten für Faserpopulationen des Hörnervs bei verschiedenen Vokalen ermittelten.

Bei geringen Pegeln antworten die Fasern auf die Signalkomponenten nahe ihrer Bestfrequenz, bei mittleren Intensitäten wird die Formantstruktur wegen eintretender Sättigung der Raten durch die Faserpopulation nicht mehr dargestellt.

Eine bessere Übertragung findet sich bei der Betrachtung der zeitlichen Struktur der Entladungsmuster. Bei geringen Intensitäten antworten die Fasern zeitgleich mit den ihrer Bestfrequenz entsprechenden Signalkomponenten. Bei höheren Lautstärken bis etwa 80 dB/SPL bleibt diese Synchronisation für den Bereich der ersten und zweiten Formanten, sowie ihrer Intermodulationsprodukte erhalten, während Komponenten außerhalb dieser Formantbereiche zeitlich nicht dargestellt werden (Sachs & Young, 1979). Sinex & Geisler (1983), die synthetische Silben (/BA/, /DA/, /GA/) als Stimuli verwendeten, kommen in einer vergleichbaren Untersuchung zu ähnlichen Resultaten. Sie stellen fest, daß die Synchronisation der Antwort eine genaue Aufzeichnung der Verläufe von Formanten und/oder Transienten liefern kann, vermuten aber, daß auch die Antwortrate, besonders bei Fasern niedriger Bestfrequenz möglicherweise zur Übertragung der (wenig energiereichen) CV-Übergänge beiträgt.

Für synthetische CV-Silbenannäherungen (/MA/, /BA/), wie natürliche Silben durch schnelle Änderungen im Kurzzeitspektrum gekennzeichnet, fand Delgutte (1980) Effekte, die nichtsimultanen Verdeckungsvorgängen (s.u.) entsprechen, wie sie Harris & Dallos (1979) für nichtsprachliche Signale beschrieben haben. Falls die Faser auf den ersten Silbenteil antwortet, ist die Reaktion auf den zweiten abgeschwächt.

Bezüglich des Einflusses der Geschwindigkeit spektraler Änderungen, einer wichtigen Eigenschaft von CV-Übergängen, auf das Antwortverhalten von Nervenzellen sind die Befunde von Möller (1974) im Cochlearis-Kern der Ratte von Interesse. Töne und Schmalbandrauschen wurden mit verschiedenen Geschwindigkeiten über einen Frequenzbereich von 3 bis 30 kHz moduliert. Für genügend hohe Modulationsraten ergab sich bei beiden Signaltypen, daß die Antwort im Bereich der Bestfrequenz schärfer und ausgeprägter wird. Einige Neurone reagieren in dieser Weise für bestimmte Geschwindigkeitsbereiche der Frequenzänderung. Nach Möller können so schnelle spektrale Änderungen im Signal betont, langsame unterdrückt und die Frequenzselektivität, d.h. die Auflösung der im Signal enthaltenen Frequenzkomponenten, erhöht werden.

Im Corpus geniculatum mediale (CGM) von Katzen stellten David et al. (1977) eine Abhängigkeit der Zeitstruktur der Zellantworten von Richtung und Geschwindigkeit der Frequenzmodulation von Signalen fest. Außerdem berichten sie von Neuronen, die auf einen komplexen tonalen Reiz, bestehend aus einem 1 kHz-Ton und einem Frequenzabfall von 10 auf 2 kHz, nicht aber auf die einzeln dargebotenen Komponenten antworteten.

Mit der Verarbeitung dynamischer Komponenten komplexer Signale in CGM und Cortex des Meerschweinchens befaßten sich Creutzfeldt et al. (1980). Sie bestimmten mittels Kreuzkorrelation und Merkmalen der Zellantworten, wie z.B. der Bestfrequenzen und der Schwellen für verschiedene Frequenzen, Neurone dieser Gebiete, die mit hoher Wahrscheinlichkeit miteinander synaptisch verbunden sind.

Als Reize dienten amplituden- und frequenzmodulierte Töne bei Bestfrequenz und natürliche Vokalisationen von Meerschweinchen.

Für die tonalen Stimuli zeigte sich, daß die Neurone des CGM noch bei höheren Modulationsfrequenzen, also schnelleren Änderungen, die Zeitstruktur der Signale in Form von Entladungsmaxima übertragen, als die nachgeschalteten cortikalen Zellen.

Die Resultate für die natürlichen Laute gehen in dieselbe Richtung. Während die CGM-Neurone die Zeit- und Amplitudenstruktur des Signals recht detailliert wiedergeben, markieren die cortikalen Zellen hauptsächlich Übergänge zwischen Signalteilen wie sie z.B. am Anfang und Ende der 'Silben' der Tiervokalisationen auftreten.

Derartige Akzentuierungen bestimmter Merkmale komplexer Signale finden sich in erhöhtem Maße in den übergeordneten Stationen der Hörbahn.

Auch die Begrenzung des Dynamikbereichs in der Peripherie wird augenscheinlich hier durch interaktive Prozesse, wie z.B. laterale Inhibition, kompensiert (vgl. Evans, 1975, 1981 b).

Für die Übertragung von Konsonant-Vokalübergängen sind demnach die Eigenschaften höherer Kerngebiete besonders angemessen. Im Sinne einer möglichst vollständigen Übertragung der Signale liegt für diese Untersuchung die Zellableitung im CGM nahe, da mit einer guten Darstellung der Konsonant-Vokalübergänge und auch der Vokale gerechnet werden kann, während die cortikalen Strukturen (vielleicht auch wegen ihrer Empfindlichkeit gegenüber Anästhesie) wohl wenig auf die eher kontinuierlichen Signalteile ansprechen.

3. Zu prüfende Annahmen

Es werden diejenigen Parameter klassischer Verdeckungsexperimente variiert, für die dies auch bei Sprachreizen sinnvoll und möglich ist.

So werden bei geeigneten Wörtern verschieden lange Intervalle (Verzögerungszeiten) zwischen Frikative (Maskierer) und Verschußlaute (Testreize) eingefügt.

Außerdem werden die Pegel der (isolierbaren) Frikative variiert, um eine Entsprechung zur Untersuchung des Einflusses verschiedener Maskiererpegel zu erhalten.

Im psychophysischen Versuch werden abweichend von Verdeckungsexperimenten keine Mithorschwellen gemessen, sondern jeweils die Verzögerungszeit, die notwendig ist, einen (überschwelligen) Testreiz, hier den Verschußlaut, wahrzunehmen.

Die Ergebnisse sollten, wenn sie eine Verdeckungsannahme schon nicht direkt beweisen können, ihr zumindest qualitativ entsprechen. Insbesondere sollten höhere Pegel der Frikative, die einen akustischen, nicht sprachgebundenen Parameter darstellen, längere Verzögerungszeiten ergeben.

Die Ergebnisse der neurophysiologischen Messungen sollten denen für Verdeckungsexperimente mit einfachen Signalen entsprechen : die Hypothese ist also, daß die Raten der Antwort auf den Testreiz, d.h. auf den dem Verschußlaut zuzuordnenden Wortteil, umso geringer ausfallen, je kürzer die Verzögerungszeiten und je höher die Maskiererpegel sind.

Bei der Einschätzung der Vergleichbarkeit psychoakustischer und neurophysiologischer Daten sind vorab einige Einschränkungen zu machen.

Wird im Verhaltensexperiment die Antwort eines Gesamtsystems erhoben (beeinflusst durch Variablen wie z.B. individuelle oder durch das experimentelle Design bedingte Entscheidungs- und Antwortstrategien), so sind neurophysiologische Daten Antworten kleiner, im Fall der Einzelzellableitung kleinster funktioneller Untereinheiten des Systems.

Dem wachen menschlichen Gehör steht im neurophysiologischen Versuch das Neuron der betäubten Katze gegenüber: Mittelohrmuskelaktivität und Einflüsse efferenter Systeme entfallen.

Darüberhinaus sind die Meßmethoden verschieden und ergeben daher Daten, die nicht ineinander überführbar sind: In den hier durchzuführenden psychophysischen Experimenten wird geprüft, ob eine bestimmte Wahrnehmung vorliegt oder nicht. Das neurophysiologische Experiment ergibt hingegen quantitative Maße der Reaktivität von Nervenzellen. Ein Kriterium für das Überschreiten einer Wahrnehmungsschwelle gibt es hier ohne begleitende Verhaltensprüfung nicht.

Diese Sachverhalte erschweren quantitative Vergleiche gemessener Daten. Ergibt sich aus psychoakustischen und neurophysiologischen Versuchen jedoch eine durchgängige qualitative Übereinstimmung mit der Annahme von Verdeckungsprozessen und der Resultate untereinander, so wird zu folgern sein, daß auditorische Maskierung bei der 'Unterdrückung' von Verschlusslauten eine Rolle spielt, also als psychoakustische Erklärung des Phänomens naheliegt.

4. Stimuli und Methoden

4.1 Auswahl der Versuchswörter

Der angestrebte Vergleich neurophysiologischer und psychoakustischer Messungen wird anhand natürlicher Äußerungen der deutschen Wörter /SPOTT/ und /STAHL/ durchgeführt. /SPOTT/ ist dem Freiburger Sprachverständlichkeitstest (DIN 45621) entnommen, /STAHL/ von mir selbst gesprochen.

Diese Beschränkung der Zahl der Versuchswörter wird durch den bei den neurophysiologischen Experimenten erforderlichen hohen Zeitaufwand notwendig. Folgende Gründe hierfür seien genannt:

- Die Stimuli sind auf ein Spektrum von 100 Hz bis 7 kHz beschränkt. Der Hörbereich der Katze reicht bis ca. 50 kHz bei einer Bevorzugung der höheren Frequenzbereiche. Der Anteil von Zellen, die auf im Sprachreiz enthaltene Frequenzkomponenten reagieren, ist entsprechend gering.

- Um sicherzugehen, daß die Zelle auf die interessierenden Wortteile antwortet, sind zusätzliche Stimulusvarianten (s.u.) notwendig.

- Es sind nicht immer alle Stimuli testbar : aufgrund der Elastizität des Hirngewebes verschiebt sich dieses gelegentlich relativ zur Position der Elektroden Spitze, oder aber es kommt zu Adaptationsprozessen. Folge ist jeweils eine Abnahme oder völliges Verschwinden der jeweiligen Zellantwort. Hier-von ist besonders die zweite Versuchsreihe betroffen, in der ein zusätzlicher Parameter variiert wird.

Psychophysische Vorversuche zeigen allerdings, daß für die Konsonanten /T/ und /P/ das Entfernen des Verschlußintervalls zwischen dem frikativen /SCH/ und der folgenden CV-Silbe mit

großer Zuverlässigkeit - unabhängig von Sprecher und Vokalkontext - zur 'Verschlußlautunterdrückung' führt.

Es wird demnach angenommen, daß auch mit nur zwei Wörtern brauchbare Ergebnisse erzielt werden können.

Für die neurophysiologischen Versuche soll hierzu beitragen, daß sich die Wörter bezüglich der Parameter, die nicht Untersuchungsgegenstand sind, möglichst unterscheiden : daher verschiedene Konsonanten, Vokale und Sprecher.

4.2 Erstellung der Stimuli

4.2.1 Testwörter

Die natürlich gesprochenen, auf Band gespeicherten Wörter werden mit einer Grenzfrequenz von 7 kHz tiefpasgefiltert und mit einer Rate von 60 μ s (ca. 16 kHz) digital abgetastet und auf Magnetplatte geschrieben.

Ihre Amplituden werden relativ zum auftretenden Maximalwert normiert, die maximale Lautstärke der (energiereichen) Vokale ist für alle Wortstimuli gleich.

Diese Stimuli werden durch 'digitalen Schnitt' erzeugt. Anfangs- und Endpunkte der Intervalle zwischen Frikativen und Verschlußlauten werden über Bildschirmdarstellung und die ausgedruckten Abtastwerte ermittelt. Ihre Längen entsprechen mit 69 ms (/STAHL/) und 122 ms (/SPOTT/) den üblichen Größenordnungen (vgl. z.B. Bailey & Summerfield, 1980).

Der hintere Wortteil (die CV-Silbe) wird dann an den vorderen herangeschoben (0 ms Intervall). Unterschiedlich lange Pausen zwischen /SCH/ und den Verschlußlauten entstehen durch Einfügen der jeweils entsprechenden Anzahl von Nullwerten.

Die Stimuli enthalten Intervalle von 0 bis 60 ms, in 7,5 ms-Schritten für die psychophysischen, in 15 ms-Schritten für die neurophysiologischen Messungen.

4.2.2 Verschlußlautreize

Zur Prüfung der Hypothese auditorischer Nachverdeckung liegt nahe, die angenommenen frikativen Maskierer zusammen mit den in verschiedenen Zeitabständen folgenden isolierten Verschlußlauten zu präsentieren. Diese sind, wie bereits erwähnt, akustisch nicht von ihrer Vokalumgebung trennbar, da koartikuliert, ihre akustischen Parameter je nach Kontext also verschieden.

Der Übergang von der gerauschhaften Öffnung des Vokaltraktes zum quasi-statischen Vokal vollzieht sich bei Verschlußlauten innerhalb der ersten 15 - 30 ms eines CV-Übergangs. (Stevens & Blumstein, 1978; Bailey & Summerfield, 1980). Bei den hier gewählten Versuchswörtern ist bereits nach 18 ms stimmhafte Periodizität im Signal enthalten. Demnach wird eine Reihe von Stimuli aus den Versuchswörtern gewonnen, bei denen der Vokal 18 ms nach Beginn der CV-Silbe abgeschnitten wird. Da dies nicht an einem Nulldurchgang geschieht, wird zur Vermeidung von Abschalteteffekten (Knacken) mit einem Cosinusfenster gewichtet (Ausschwingzeit 6 ms).

Auf gleiche Weise und in entsprechender Länge wie für die vollständigen Wörter eingefügte Intervalle ergeben jeweils eine Reihe der Einfachheit halber /ST/- bzw. /SP/- Stimuli genannter Verschlußlautreize.

4.2.3 Variation der Maskiererpegel

Zur Untersuchung des Einflusses des Maskiererpegels werden die Amplituden der Frikative bei den Wort- und Verschußlautreizen variiert.

Die natürlichen /SCH/ von /STAHL/ und /SPOTT/ werden isoliert und beider Amplituden mittels Rechner so angehoben, daß ihr Spitzenwert dem der Vokale entspricht. Dieser höchste Maskiererpegel wird mit 0 dB bezeichnet. Anschließend wird dieser Pegel zweimal um je 10 dB abgesenkt, sodaß sich pro Wort drei verschieden laute Frikative (0, -10, -20 dB) ergeben. Diese werden mit den verschiedenen langen Intervallen den Konsonant - Vokalsilben bzw. -übergängen vorangestellt.

Relativ dazu haben die originalen Frikative annähernd Pegel von -24 dB (/STAHL/) und -16 dB (/SPOTT/).

4.2.4 Zusätzliche Stimuli

Wegen variabler Synchronizität (Latenz und Ausdehnung) der Antworten verschiedener Zellen wird mittels weiterer Stimuli sichergestellt, daß die Nervenzellen tatsächlich auf die den Verschußlauten entsprechenden Signalteile reagieren :

- /T/ und /P/

Jeweils 18 ms des Beginns der CV-Silben werden herausgeschnitten und mit einem Cosinusfenster von 6 ms Ausschwingzeit gewichtet. Nur Neurone, die auf diese isoliert angebotenen Reize ansprechen, werden weiter untersucht.

- /TAHL/ und /POTT/

Nach Entfernen der Frikative verbleiben die der Verschußphase folgenden Wortteile. Diesen werden Intervalle von

15, 30, und 60 ms Dauer zwischen den ersten 18 ms der Silben-
beginne und den folgenden Vokalen eingefügt, Aus- und Ein-
schwingen mit 6 ms bzw. 12 ms Cosinusfenster gewichtet. Die
je Wort vier Reize dienen der Prüfung auf mögliche Vorverdek-
kung und eventuell der Trennung von 'Konsonant- und Vokal-
Antwort', falls ein Neuron auf beide Wortteile anspricht.

- /A/ und /O/

Von den Wortteilen /TAHL/ und /POTT/ werden die ersten
18 ms abgeschnitten, die verbleibenden Vokalansätze mit einer
Einschwingzeit von 12 ms Cosinusgewichtet. Die isolierten Vo-
kale dienen der Kontrolle, ob, und wenn, in welchem Ausmaß
sich eine Zellantwort auf den Vokalteil eines Testwortes be-
zieht.

- /SCHAL/ und /SCHOTT/

Bei den Wortstimuli mit 0 und 60 ms Intervalllänge werden
die dem Einschwingen der CV-Silbe entsprechenden 18 ms ent-
fernt, die Vokalanfänge mit einem Cosinusfenster von 12 ms
Länge gewichtet. Mittels dieser Reize kann zusätzlich geprüft
werden, ob sich Unterdrückungseffekte ausschließlich auf die
den Konsonanten entsprechenden Signalteile beziehen.

- 120 ms Intervalle

Für Testwörter und Verschlusslautreize werden im
neurophysiologischen Experiment auch Intervalle von 120 ms
Länge angeboten. Die hier gemessenen Antworten auf die Ver-
schlusslautteile der Signale werden mit denen für 60 ms Pau-
senlänge verglichen, um zu klären, ob eine Unterdrückung über
den für die psychophysischen Messungen bedeutsamen Bereich
hinaus stattfindet.

- Bandpaßrauschen

Digital gespeichertes Rauschen im Frequenzbereich von 100 Hz bis 7 kHz dient zur Vorauswahl von Neuronen, die innerhalb des Spektrums der Sprachreize sensitiv sind.

4.3 Psychoakustische Messungen

4.3.1 Reizdarbietung

Die Wortstimuli werden per Digitalrechner (LSI 11/23) D/A-gewandelt, tiefpaßgefiltert (Grenzfrequenz 7 kHz) und verstärkt (Hitachi HCA/HMA 7500 Mk II) über Kopfhörer (Sennheiser HD 230) dargeboten.

Die neun verschiedenen Intervallängen von 0 - 60 ms werden je 10 mal wiederholt, die Reihenfolge der Reize wird von einer rechnergenerierten Zufallsfolge bestimmt.

Bei der ersten Versuchsreihe entscheiden die Versuchspersonen durch Eingabe der entsprechenden Antwort in einer 'Yes-No'-Prozedur, ob sie bei dem angebotenen Wort ein /T/ resp. ein /P/ hören. Auf diese Antwort folgt die Darbietung des nächsten Reizes.

Für die zweite Versuchsreihe wurde ein AXB-Design gewählt. Der zu beurteilende Stimulus befindet sich mit Abständen von jeweils einer halben Sekunde zwischen den Wortreizen mit den extremen Intervallängen.

Die Versuchsperson gibt an, ob der jeweilige mittlere dem ersten oder dem zweiten umgebenden Reiz ähnlicher klingt.

Sowohl die Position der Reize mit 0 ms (A) oder 60 ms (B) Intervall, wie auch die Reihenfolge der zu beurteilenden X-Stimuli ist zufallsvariiert.

Das AXB-Design wurde für die Messungen bei variiertem

Maskiererpegel vorgezogen, weil sich gezeigt hat, daß es bei den leisen Frikativen Schwierigkeiten gab, ein konstantes Entscheidungskriterium für eine Ja- oder Nein- Antwort zu finden. Als Folge ergaben sich hohe Anteile von Verschlußlautantworten schon bei den kürzesten Intervallen. Bei der Prüfung der AXB-Anordnung mit Stimuli der ersten Versuchsreihe wurden auffällige Abweichungen von den Ergebnissen des 'Yes-No'-Experiments nicht festgestellt, was den theoretischen Erwartungen entspricht (vgl. Macmillan et al., 1977, Pastore, 1981).

4.3.2 Datenerhebung

Im 'Yes-No'-Verfahren werden für die verschiedenen Pausenlängen jeweils die Ja-Antworten der Versuchsperson addiert, im AXB-Fall die Urteile, der X-Reiz klinge dem Wort mit dem Intervall von 60 ms ähnlicher.

Diese Summen werden am Ende der Einzelversuche ausgegeben. Später werden sie dann über die Versuchspersonen in relative Häufigkeiten, dargestellt in Prozentwerten, umgeformt.

4.4 Neurophysiologische Messungen

4.4.1 Vorbereitung der Versuchstiere

Als Versuchstiere dienen erwachsene Katzen. Die Tiere sind akut präpariert und während des Versuchs anästhetisiert (1.75 mg Pentobarbital per kg Körpergewicht und Stunde) und paralysiert (10 mg Gallamin-triethiodid/kg/h). Die Beatmung erfolgt künstlich (Harvard-Kleintier-Atempumpe).

Die notwendigen chirurgischen Eingriffe werden am narkotisierten Tier (35 mg Ketanest/kg) vorgenommen. Als erstes wird an einem Vorderbein ein intravenöser Zugang für das Anästhetikum geschaffen, anschließend ein Luftrohrenschnitt für die künstliche Beatmung ausgeführt.

Dann wird daß Versuchstier mittels zweier in die Gehörgänge eingeführter Metallstäbe in einem Rahmen positioniert, an dem ein stereotaktisches Gerät angebracht ist. Die Schädeldecke wird freigelegt und über einen Halter aus Stahl, an der Schädeldecke mit Zahnzement und kleinen Schrauben befestigt, mit dem Rahmen fest verbunden.

Senkrecht über dem zur Ableitung vorgesehenen Kerngebiet (CGM) wird mit einem Kreisbohrer der Schädelknochen in einem Durchmesser von 5 mm entfernt, um die (mehrfache) Einführung von Meßelektroden zu ermöglichen.

Während der Ableitung wird dieses Loch mit Agar, einer gallertartigen Masse, verschlossen, um der Austrocknung des Gewebes vorzubeugen.

Die Wirksamkeit der künstlichen Beatmung wird regelmäßig anhand des CO_2 -Gehaltes der ausgeatmeten Luft überprüft.

Zusätzliche Information über den Zustand des Tieres geben osziloskopische und akustische Wiedergabe des kontinuierlich abgeleiteten EKG.

Die Dauer der Versuche beträgt für ein Tier etwa 30 Stunden, auf zwei Tage verteilt. Danach wird die Katze entweder durch eine Überdosis Pentobarbital eingeschlafert, oder zwecks histologischer Bestimmung des genauen Ableitungs-ortes perfundiert.

(Perfusion: 1. Spülung des Gehirnkreislaufes mit einer Salzlösung zur Entfernung des Blutes. 2. Spülung mit Formalin zur Konservierung der Hirnstrukturen)

4.4.2 Ableitungstechnik

Der Rahmen mit dem Versuchstier ist in einer schallgedammten Kammer (Siemens) fest installiert.

Anhand von Hirnkoordinaten, die als durchschnittliche Werte über Atlanten zugänglich sind, erlaubt das stereotaktische Gerät eine genaue Plazierung der Meßelektrode senkrecht über dem gewünschten Kerngebiet. Ihre vertikale Positionierung wird mit einem hydraulischen Mikromanipulator verändert.

Bei den Elektroden handelt es sich entweder um mit leiterlösender Lösung gefüllte Glas-Mikropipetten oder um elektrolytisch angespitzte Karbonfasern. Ihr Durchmesser an der Spitze beträgt etwa 1 μm (Glas) bzw. 7 μm (Karbonfaser). Gemessen bei 1000 Hz Wechselstrom in physiologischer Ringerlösung weisen sie Widerstände zwischen 4 und 12 M Ω auf.

Die extrazellulär abgeleiteten Ströme werden innerhalb der Kammer vorverstärkt, durchlaufen außerhalb einen weiteren Verstärker, werden dann bandpaßgefiltert (0.1 - 1 kHz) und auf einem Oszilloskop dargestellt.

Durch einen variablen Schmitt-Trigger bzw. einen Fensterdiskriminator können Aktionspotentiale (Spikes) einzelner Nervenzellen selektiert werden, indem nur Spannungen gewählter Größe weitergeleitet werden. Diese werden in Einheitsimpulse vom 0.2 ms Dauer und 5 V Amplitude umgewandelt, welche dann per Digitalrechner (PDP 11/40) on-line zu einem reizsynchronen Peri-Stimulus-Time Histogramm (PSTH, wie z.B. in Abb. 5, S. 39) verarbeitet werden. Die Zellantworten werden über jeweils 50 Reizdarbietungen mit einer zeitlichen Auflösung (Binweite) von 1 ms aufaddiert.

Die Auswertung der gespeicherten Meßdaten erfolgt off-line ebenfalls mit Hilfe der Rechenanlage.

4.4.3 Reizdarbietung

In der ersten neurophysiologischen Versuchsreihe werden die Stimuli über einen etwa einen Meter vom Versuchstier entfernten aktiven Lautsprecher (Philips MFB 587), also binaural, in der zweiten monaural über die contralaterale Muschel eines Kopfhörers (Sennheiser HD 414) je 50 mal dargeboten.

Zwischen den Reizwiederholungen befinden sich Pausen von einer Sekunde.

Die Lautstärke der Darbietung beträgt etwa 40 dB über der menschlichen Schwelle für einen Ton von 1 kHz.

4.4.4 Ermittlung der Antwortraten

Die im neurophysiologischen Experiment gewonnenen Daten - auf Disketten gespeichert - können auf dem Bildschirm des Laborrechners (LPS-11) als PSTH dargestellt werden. Zusätzlich wird im oberen Teil des Display zeitsynchron der jeweilige Stimulus bzw. sein Amplitudenverlauf halbgleichgerichtet abgebildet.

Über den gewünschten Histogrammteil kann ein Fenster variabler Länge gelegt, die Summe der innerhalb dieses Fensters gezählten Spikes abgelesen werden.

Anhand der für den isolierten Konsonant-Vokalübergang gewonnenen Zellantworten werden die jeweilige Antwortlatenz (es kommen in dieser Untersuchung fast nur kurze Latenzen von 9 bis 12 ms vor) und die sinnvolle Weite des Fensters (10 bis 18 ms) bestimmt.

Dann werden die Spikesummen an den dem CV-Übergang entsprechenden Wortstellen (d.h. innerhalb dieser Fenster) für die verschiedenen langen Intervalle ermittelt.

Da die Aktivität der einzelnen Neurone zeitlichen Schwankungen unterliegt, im Interesse der Vergleichbarkeit der verschiedenen Histogramme einer Meßreihe aber konstant sein sollte, empfiehlt sich eine diesbezügliche Korrektur. Diese wird in der Regel über die Spontanaktivität der Neurone vorgenommen. Deren Messung mußte hier wegen der langen Zeit, die für die Darbietung einer vollständigen Reihe von Stimuli anfällt, allerdings mehrfach wiederholt werden.

Unter mehreren getesteten Möglichkeiten für einen Bezugs- wert erweist sich diejenige Antwortaktivität, die vom Beginn der Reizdarbietung bis zum Ende des frikativen /SCH/ auftritt, als sinnvoll. Sie enthält etwa 150 ms Spontanaktivität und die Antwort auf den Wortteil, dessen Einfluß auf folgende Signalkomponenten untersucht wird; sie ist außerdem recht stabil, sodaß sich keine extremen Korrekturen der Antwort auf den /T/- bzw. /P/-Anteil des Signals ergeben.

Es werden demnach zusätzlich die Summen der Spikes für die ersten 350 ms (/STAHL/, /ST/) und 380 ms (/SPOTT/, /SP/) der Zellantworten erhoben. Diese werden über die verschiedenen Intervalle gemittelt, diese Mittelwerte dann durch die Spikesummen für die einzelnen Intervalle dividiert. Die so für jedes Histogramm erhaltenen Korrekturfaktoren werden mit der für den 'Konsonanten' abgelesenen Rate multipliziert.

Die Korrekturfaktoren bewegen sich mit seltenen Ausnahmen zwischen 0.7 und 1.3 .

5. Besprechung der Ergebnisse

5.1 Psychoakustische Messungen

5.1.1 Erste Versuchsreihe

Die Ergebnisse der 'Yes - No' Messungen sind für 6 teilweise 'naive' Versuchspersonen in Abb. 3 als Übergänge zwischen /SCHAL/ und /STAHL/ bzw. /SCHOTT/ und /SPOTT/ dargestellt.

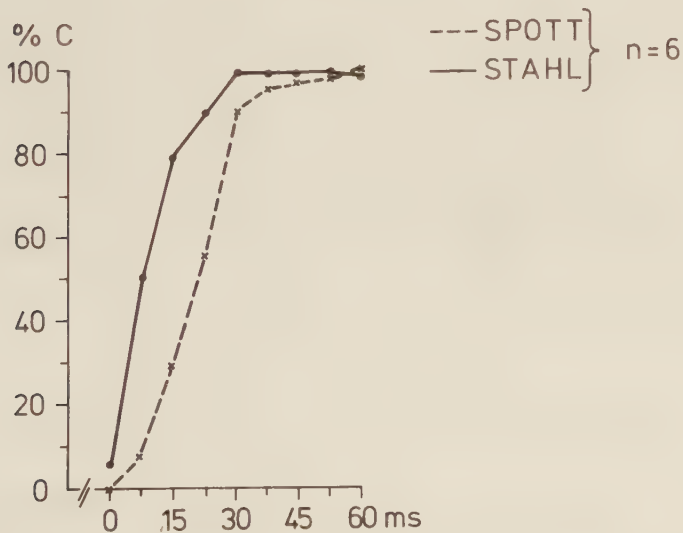


Abb. 3 : Mittlere relative Häufigkeiten der Ja - Urteile für die Wahrnehmung eines Verschlußblautes (C).

Eine Schätzung der kritischen Pausenlangen mittels einer Regressionsanalyse (Sixtl, 1967, Sachs, 1972) ergibt ausgehend von den von 0 und 100 % verschiedenen mittleren relativen Urteilen 10.2 ms für /STAHL/ und 20.1 ms für /SPOTT/. Mit jeweils 93 % durch Linearität erklärter Varianz ist diese Annäherung den Daten gut angepaßt. Somit liegen die gemessenen Übergangszeiten innerhalb der Größenordnungen, die von anderen Autoren gefunden wurden. Dorman et al. (1979) geben für ihre synthetischen Stimuli mit notwendigen Längen des Intervalls von etwa 40 ms bei Verschlußlauten am Silbenbeginn und 58 ms bei postvokalischer Stellung etwas höhere Werte an, ihre Werte für natürliche Sprachsignale fallen um etwa 15 ms geringer aus. Die Daten von Best et al. (1981) zeigen Übergänge bei etwa 25 und 55 ms, während Bailey & Summerfield (1980) Verschlußlautwahrnehmungen bei Intervallen von 10 - 40 ms Länge mit synthetischen 3-Formantstimuli finden, wobei bilabiale Laute bevorzugt bei längeren Pausen gehört werden. Eine Untersuchung mit natürlichen Äußerungen der englischen Wörter /SPIN/ und /SKIN/ (Spinner & Schreiner, 1980) ergibt Zeiten zwischen 15 und 30 ms, dabei wiederum längere für /P/.

5.1.2 Zweite Versuchsreihe

Die Resultate im AXB-Experiment mit je drei verschiedenen Pegeln der Frikative zeigt die Abb. 4 für zwei geübte Versuchspersonen.

Der Übersichtlichkeit halber sind die Kurven nur bis zu jenem Intervall aufgetragen, bei dem erstmals in 100 % der Fälle der Verschlußlaut gehört wird.

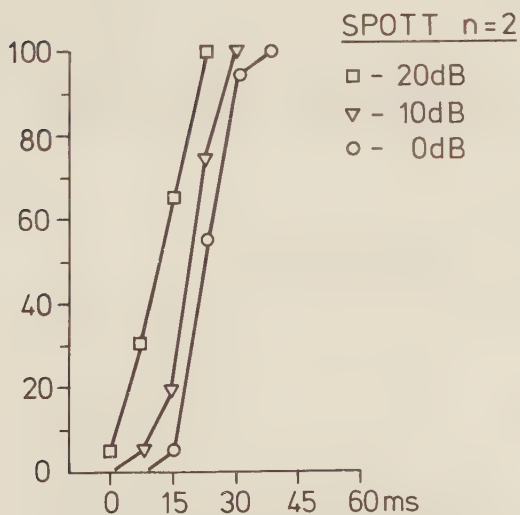
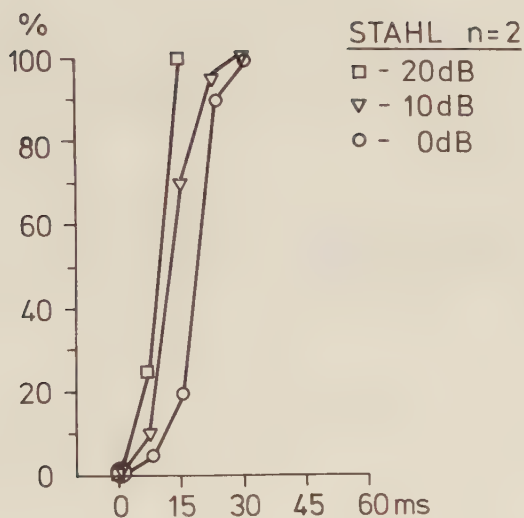


Abb. 4 : Mittlere relative Urteilshäufigkeiten ($X = B$) für drei Maskiererpegel und zwei Vpn. über Intervallängen.

Man sieht, daß die gemessenen Intervallängen mit wachsendem Maskiererpegel größer werden. Die Überprüfung mittels des verteilungsfreien Trend-Tests von Page (nach Lienert, 1975) ergibt, daß Pegel und Urteilshäufigkeiten für Intervalle bis 15 ms (/STAHL/) bzw. 22.5 ms (/SPOTT/) gleichsinnig geordnet sind. Diese Zunahme der Verzögerungszeiten ist signifikant ($p < 0.05$).

Die Länge des zur Wahrnehmung der Verschlußlaute notwendigen Intervalls wird demnach vom Pegel der vorangehenden Frikative, einem akustischen Parameter, beeinflusst.

Diesbezügliche Schätzungen mittels Regressionsanalyse ergeben für die Pegel 0, -10 und -20 dB kritische Pausenlängen von 8.1, 14.1 und 16.2 ms bei /STAHL/, 9.6, 16.6, sowie 22.6 ms bei /SPOTT/, also wie in der ersten Versuchsreihe jeweils längere für /P/. Die Linearitätsannahme erklärt wiederum jeweils mehr als 93 % der Varianz der mittleren Urteile.

Der gezeigte Einfluß des Pegels der Frikative unterstützt die Annahme von Nachverdeckungsvorgängen.

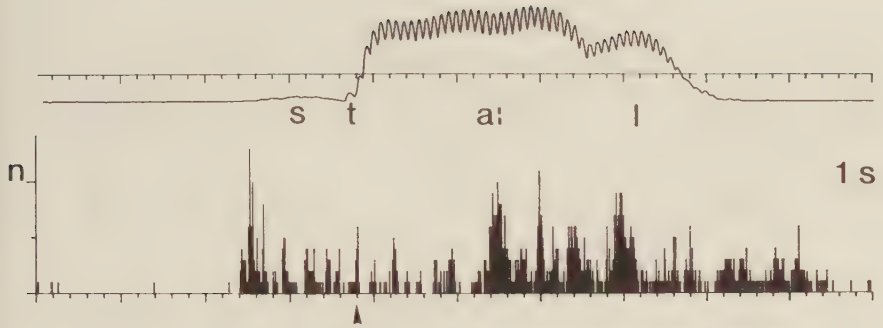
5.2 Neurophysiologische Messungen

5.2.1 Erste Versuchsreihe

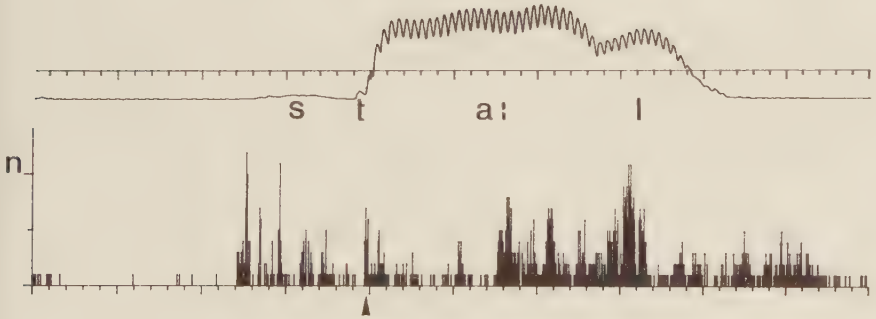
In der ersten Versuchsreihe wurde von insgesamt 34 Zellen des CGM abgeleitet. Dabei antwortete ein Teil der Neurone nur auf bestimmte, einige auf alle Stimuli. Auch konnten einzelne Zellen nicht lange genug abgeleitet werden, um über jeweils alle Intervallängen Messungen zu ermöglichen.

Die Spontanaktivität der Neurone ist mit wenigen Ausnahmen sehr gering.

STAHL
15



30



60ms

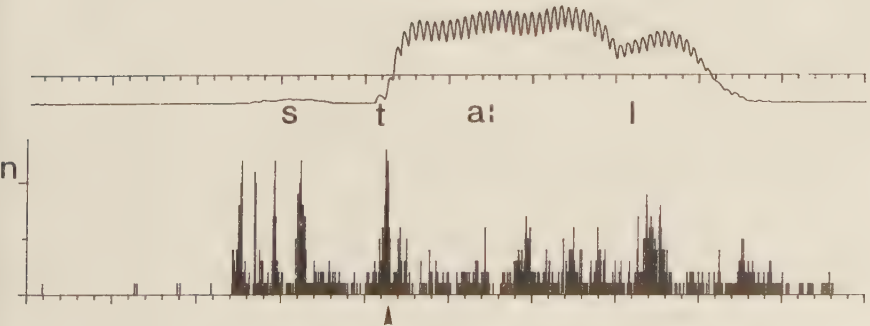


Abb. 5 : Von einem Neuron abgeleitete Peri-Stimulus-Time Histogramme für das Wort /STAHL/ mit Intervallen von 15, 30, und 60 ms. Jeweils oberhalb das Zeitsignal.

Abb. 5 zeigt von einem Neuron abgeleitete Histogramme und jeweils oberhalb die synchrone Aufzeichnung des Amplitudenverlaufs des Zeitsignals für /STAHL/. Die zur Darstellung ausgewählten Intervalllängen betragen 15, 30 und 60 ms. Die dem /T/ zugeordneten Zellantworten sind durch Pfeile markiert. Man sieht eine deutliche Zunahme der Entladungsrate mit wachsender Pausendauer.

Bei diesem Neuron ist im Gegensatz zu anderen auch die Antwort auf die anderen Wortteile recht deutlich ausgeprägt, die Spontanaktivität ist höher als die Regel.

Beim Vergleich mit Wahrnehmungsdaten als Produkt ganzer Sinnessysteme ist es sinnvoll, sich nicht auf Antworten einzelner Nervenzellen, sondern auf die einer größeren Anzahl zu beziehen. Zur Darstellung einer solchen Populationsantwort werden die korrigierten Antwortraten über die abgeleiteten Zellen für jedes Intervall arithmetisch gemittelt. Als alternatives Maß kommt der (häufig gewählte) Median der Zellantworten in Frage. Bei den vorliegenden Daten ergeben sich zum Mittelwert geringe Unterschiede im Sinne einer Betonung der Zunahme der Raten mit wachsendem Intervall. Das arithmetische Mittel ist auf die Verdeckungsannahme bezogen also hier ein konservativerer Kennwert.

Die Mittelwerte sind für die vier Stimulusgruppen in Abb. 6 aufgetragen. Es sind alle Meßpunkte berücksichtigt, also auch solche, die aus unvollständigen Meßreihen stammen.

Es zeigen sich entsprechend der Annahme von Nachverdeckungsprozessen mit längeren Verzögerungszeiten zunehmende Antwortraten bei Wort- und Verschlusslautreizen.

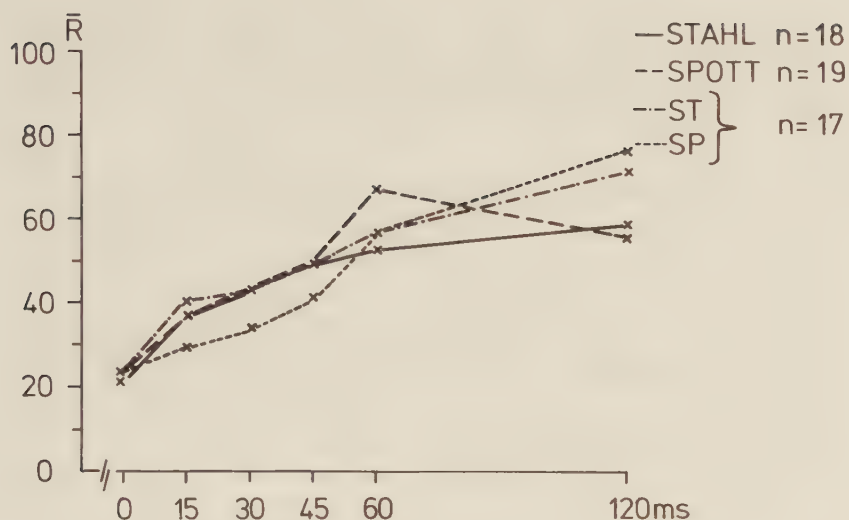


Abb. 6 : Mittlere korrigierte Antwortraten für Wort- und Verschlusslautreize.

Ange-sichts der grundsätzlichen Verschiedenheit der Maße im psychoakustischen und im neurophysiologischen Experiment (relative Urteile vs. Spikesummen) ist eine Prüfung der Hypothese qualitativer Übereinstimmung angemessen.

Deshalb, aber auch wegen der großen Unterschiede des Antwortverhaltens zwischen Zellen (s.u.) werden zur statistischen Auswertung der Ergebnisse (verteilungsfreie) Rangverfahren gewählt (Lienert, 1975).

In die Prüfung der Verdeckungsannahmen gehen im Folgenden immer nur die Antworten der einzelnen Zellen ein, bei denen über den betrachteten Bereich von Pausenlängen vollständig abgeleitet werden konnte.

Da im zum Vergleich anstehenden psychoakustischen Experiment nur Pausenlängen bis 60 ms angeboten wurden, wird als erstes untersucht, ob sich die Antwortraten für 120 ms von denen für 60 ms unterscheiden.

Nach Abb. 6 fallen augenscheinlich die mit den Verschlußlautreizen gemessenen Antwortraten für 120 ms Intervalllänge gegenüber 60 ms im Gegensatz zu denen bei den Wortreizen höher aus. Möglicherweise treten demnach bei /ST/ und /SP/ eventuelle Nachverdeckungseffekte über den psychoakustisch relevanten Bereich hinaus auf.

Also werden für Wort- und Verschlußlautreize die Antwortraten bei 60 ms mit denen bei 120 ms verglichen. Abb. 7 zeigt für beide Testwörter die mittleren korrigierten Raten über die Zellen, die jeweils bei beiden Stimulusgruppen vollständige Meßreihen lieferten.

Zur statistische Prüfung auf Unterschiede ist der Wilcoxon - Rangsummentest geeignet. Bei einseitiger Fragestellung zeigt sich bei den Wortreizen keine signifikante Zunahme der Antwortraten über 60 ms Verzögerungszeit hinaus, bei /SP/ ist die Zunahme statistisch bedeutsam ($p < 0.05$), bei /ST/ findet sich ein entsprechender Trend ($p = 0.08$).

Eine mögliche Erklärung ist Vorverdeckung des Verschlußlautanteiles durch den energiereichen Vokalteil der Wortstimuli.

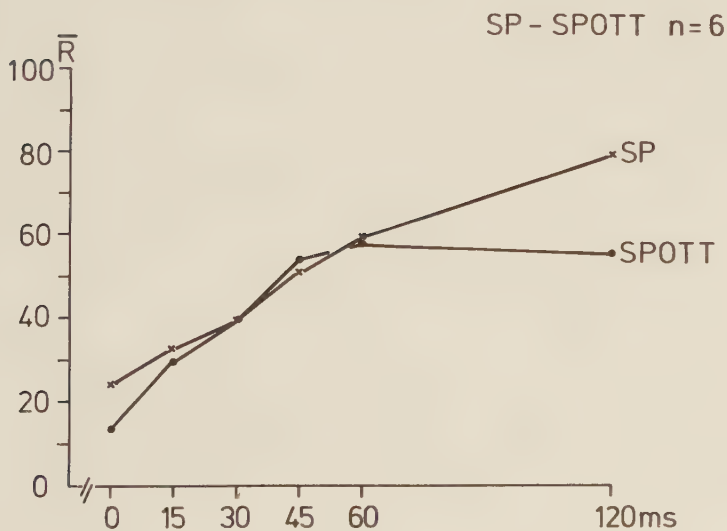
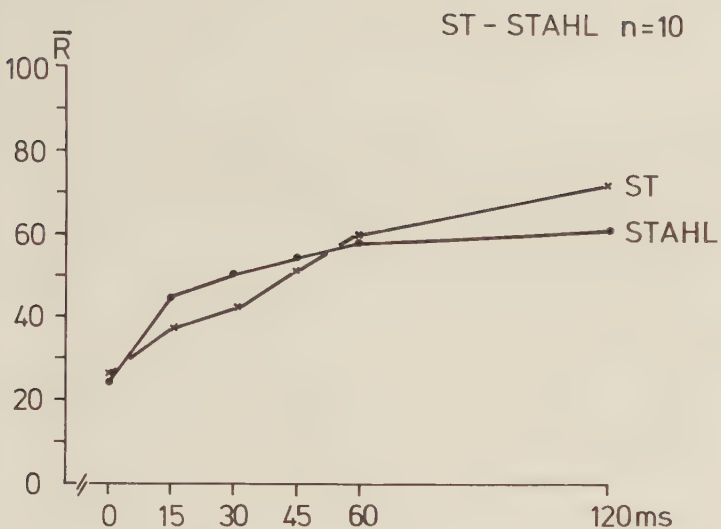


Abb. 7 : Vergleich der Antwortraten bei Wort- und Verschl  lautreizen

Die Dauer der CV-Übergänge liegt mit 18 ms innerhalb des als wirksam beschriebenen Bereichs von etwa 20 ms (Zwicker, 1982, Harris & Dallos, 1979). Trifft diese Erklärung zu, so sollten die Antwortraten für die Verschlußlautreize allgemein größer sein. Mit dem Wilcoxon-Test finden sich signifikant höhere Werte beim Paar /SP/ - /SPOTT/ nur bei den Pausenlängen 0 ms und 120 ms ($p < 0.05$), lediglich ein Trend ergibt sich bei /ST/ - /STAHL/ (120 ms, $p < 0.1$). Die Annahme von Vorverdeckungsprozessen bestätigt sich also hier nicht, wenn auch für einzelne Neurone anhand der dazu vorbereiteten Stimuli (/TAHL/, /POTT/) entsprechende Meßergebnisse vorliegen : Hier sind die dem Verschlußlautanteil zugeordneten Raten dann höher, wenn der Vokalteil in größerem Abstand folgt.

Stattdessen ist aus dieser Prüfung zu schließen, daß für alle Stimulusgruppen bis zu Intervallen von 60 ms Länge die jeweiligen Raten verwandter Größenordnung, also miteinander vergleichbar sind.

Zur statistischen Beurteilung der zentralen Frage, ob Nachverdeckungseffekte vorliegen, bietet sich wiederum der Trend - Test von Page an. In diesem Falle wird geprüft, ob längere Verzögerungszeiten höhere Rangplätze der Antwortraten zur Folge haben.

Da sich für die Wortreize keine bedeutsamen Unterschiede zwischen den beiden längsten Intervallen ergeben, gehen in Entsprechung zu den psychoakustischen Messungen die Zellen in den Test ein, für die Antwortraten bei allen Pausenlängen bis 60 ms vorliegen. Diese Beschränkung ist auch bezüglich der Verschlußlautreize sinnvoll, weil sie zu einer konservativeren Prüfung im Sinne der Nullhypothese führt.

Bei jeder der vier Stimulusgruppen findet sich ein sehr signifikanter ($p < 0.01$) Anstieg der Antwortraten bei wachsender Intervalllänge. Dies bedeutet eine gesicherte

qualitative Übereinstimmung mit dem Verhalten von Zellantworten in Nachverdeckungsexperimenten.

Quantitativ werden diese Nachverdeckungseffekte sinnvollerweise als Erholungskurven in Prozentwerten dargestellt.

Die maximale Antwort eines jeden Neurons erhält dazu den Wert 100 %, die Antwortraten für die anderen Intervalle werden entsprechend umgeformt. Die so erhaltenen relativen Antworten der Einzelzellen werden zur Darstellung einer Populationsantwort über die Verzögerungszeiten gemittelt.

Abb. 8 zeigt diese Erholungskurven für Wortreize und Intervalle bis 60 ms, bei den Verschlusslautreizen (Abb. 8 unten) gehen wegen der statistisch bedeutsamen weiteren Zunahme auch die Raten für 120 ms Pausenlänge in die Darstellung ein.

Dabei entsprechen die Antworten beim kürzesten Intervall mit etwa 30 % der erhaltenen Antwort den Angaben von Harris & Dallos (1979), die den Einfluß der Verzögerungszeit auf Antwortraten im Hörnerv für ebenfalls überschwellige Maskier- und Testreize bei Bestfrequenz untersuchten, wobei der Maskiererpegel 10 dB über dem des Testtons lag. Auch erscheinen die Größenordnungen der relativen Raten über die verschiedenen Verzögerungszeiten als vergleichbar.

Die Verläufe der Erholungskurven sind hier mit guter Anpassung linear anzunähern. Die Regression der zu z-Werten transformierten relativen mittleren Raten gegen die Intervalllänge ergibt Schätzungen für die durch die lineare Komponente erklärte Varianz von jeweils über 94 %.

Harris & Dallos favorisieren exponentielle Darstellungen, auch wegen der von ihnen vorgeschlagenen Erklärung von Nachverdeckung als Adaptationsphänomen. Davon abweichend zeigen die im CGM mit den komplexen Sprachstimuli gemessenen Ergebnisse keine Abhängigkeit der Verdeckungseffekte von der

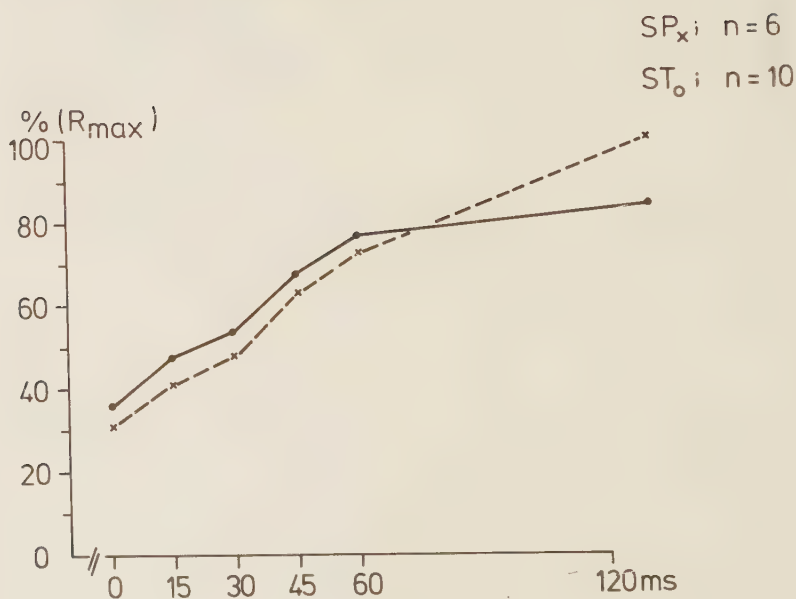
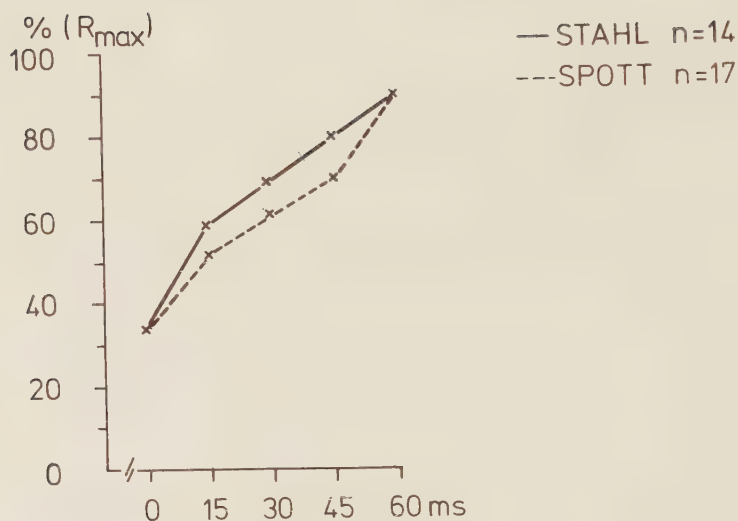


Abb. 8 : Relative Antwortraten (Erholungskurven) bei Wortreizen bis 60 ms, bei Verschlusslautreizen bis 120 ms Intervalllänge.

Antwort auf den Maskierer. Selbst wenn diese völlig fehlt, finden sich mit größerem Intervall wachsende Antwortraten.

5.2.2 Zweite Versuchsreihe

In der zweiten Versuchsreihe wurde von insgesamt 15 Einzelzellen abgeleitet.

Um eine sinnvolle Auswertung über alle Intervalle bei den je drei Maskiererpegeln zu gewährleisten, wurden nur solche Neurone untersucht, die mit einer genügend hohen Rate auf die Verschlusslautreize antworteten. Deswegen und aufgrund der für vollständige Meßreihen deutlich verlängerten Zeit fallen die Stichprobenumfänge gegenüber dem ersten neurophysiologischen Experiment geringer aus.

So konnten für /ST/ an neun, für /STAHL/ und /SP/ an je vier Zellen über alle 18 Stimuli Antwortraten ermittelt werden.

Bei /SPOTT/ ist die Anzahl der Zellen, die verwertbare Antworten liefern (zwei), zu gering, um sie in die Darstellung und Auswertung der Ergebnisse sinnvoll einzubeziehen.

Da jedoch im psychoakustisch relevanten Bereich von Pausenlängen die Ergebnisse der ersten Versuchsreihe bei den vier Stimulusgruppen wenig voneinander abweichen, sollte anhand der hier verbleibenden drei das Verhalten der Antwortraten bei verschiedenen Maskiererpegeln in ausreichendem Umfang darzustellen sein.

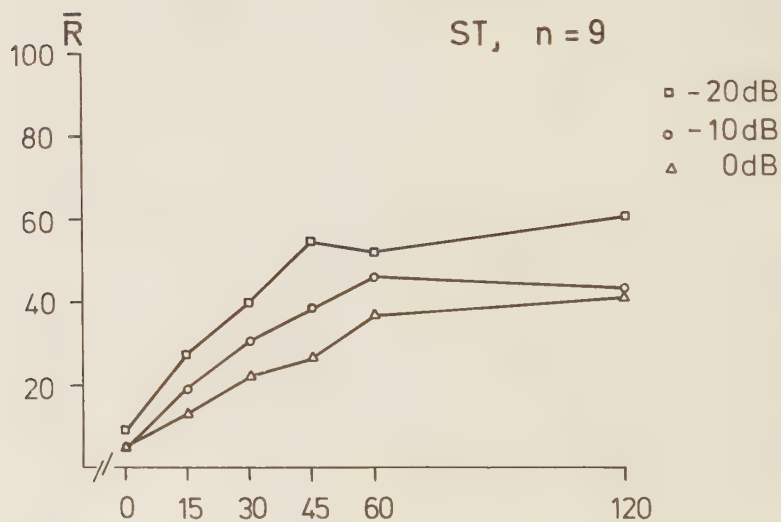
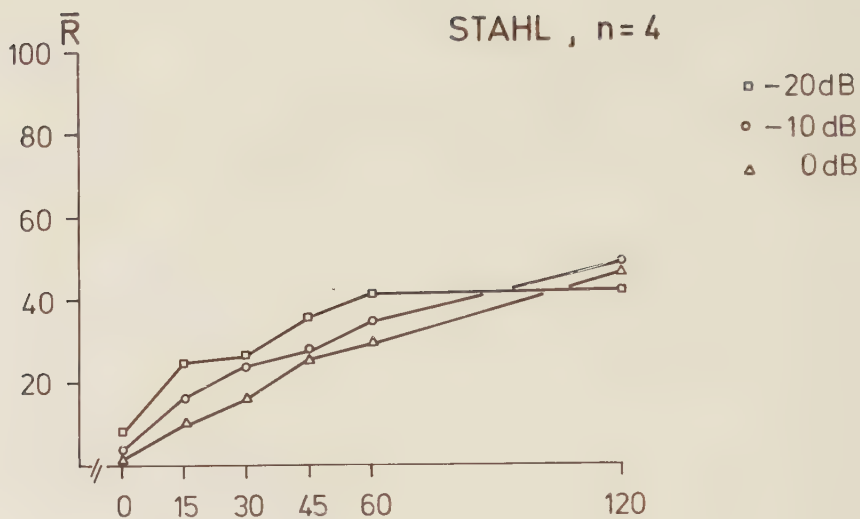


Abb. 9 : Mittlere Antwortraten bei /STAHL/ und /ST/ für drei Pegel der Frikative.

Die Abb. 9 zeigt die mittleren korrigierten Raten für /STAHL/ und /ST/. Parameter ist der maximale Pegel des frikativen /SCH/ mit 0, 10 und 20 dB unter dem Spitzenwert der Wortreize, also der Vokale.

Wie auch in der ersten Versuchreihe nehmen die Mittelwerte der Raten zumindest bis zu Intervallängen von 60 ms zu. Der Einfluß des Maskiererpegels zeigt sich als Verlangsamung dieser Zunahme bzw. in niedrigeren Raten bei höherem Pegel.

Großere Variabilität findet sich in den Ergebnissen für /SP/ (Abb. 10), wo es wegen extremer Antworten eines der Neurone zu Überlappungen der Kurven verschiedener Pegel bei den längeren Intervallen kommt.

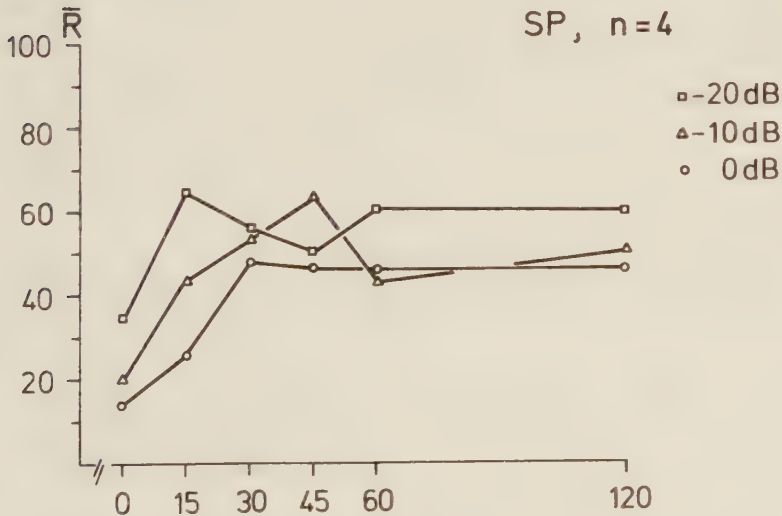


Abb. 10 : Mittlere Antwortraten bei /SP/
für drei Pegel der Frikative

Die Prüfung der Verdeckungsannahme über Intervalle bis 60 ms mittels des Trend-Tests von Page ergibt bei /STAHL/ und /ST/ einen sehr signifikanten Einfluß ($p < 0.01$) der Verzögerungszeiten bei allen Maskiererpegeln. Bei /STAHL/ ergeben sich vereinzelt im Betrag gleiche Antwortraten (Null) für benachbarte kurze Intervalle. Diesen wurden Rangplätze im Sinne der Nullhypothese zugeordnet.

Bei /SP/ ist die Zunahme der Antwortraten mit wachsenden Pausenlängen beim niedrigsten Pegel von -20 dB signifikant ($p < 0.05$), bei den beiden höheren ebenfalls mit einer Irrtumswahrscheinlichkeit von $p < 0.01$.

Auch die Wirkung der Pegelvariation wird mit dem Trend-Test geprüft. Die Verdeckungsvorgänge zugrundelegende Hypothese besagt, daß die Antwortraten mit wachsendem Pegel der Frikative jeweils geringer ausfallen.

Mit Irrtumswahrscheinlichkeiten von $p < 0.05$ oder $p < 0.01$ führen höhere Pegel zu geringeren Raten, bei /STAHL/ bis zu Pausen von 45 ms, bei /ST/ von 60 ms Länge.

Für /SP/ ist dieser Pegeleinfluß bei den Intervallen bis 30 ms signifikant.

Bei 120 ms Verzögerungszeit finden sich in keinem Fall bedeutsame Unterschiede zwischen den Antwortraten für verschiedene Pegel.

Die im ersten Versuch festgestellte Abhängigkeit der Antwortraten von der Verzögerungszeit und damit die qualitative Entsprechung zu Ergebnissen in Nachverdeckungsexperimenten wird also durchgehend bestätigt.

Die Annahme von Verdeckungsvorgängen wird hier zusätzlich begründet durch Entsprechungen, die die Wirkung verschiedener Maskiererpegel auf die Reduktion der Antwortraten betreffen.

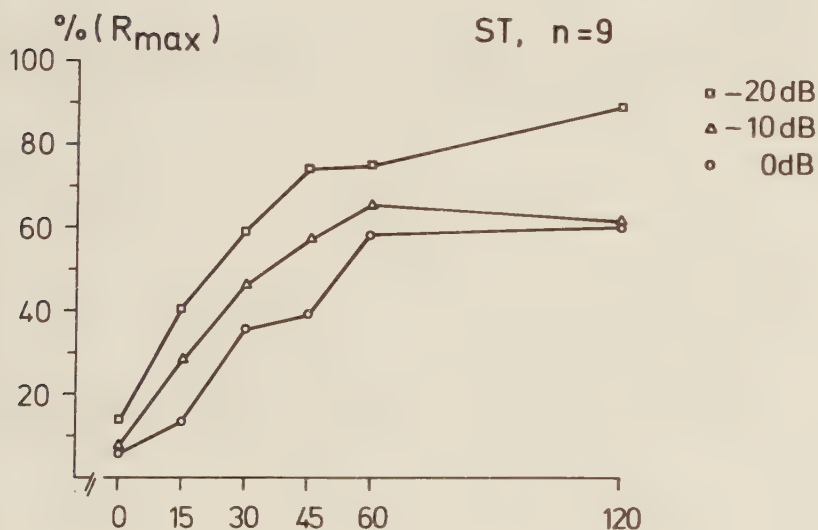
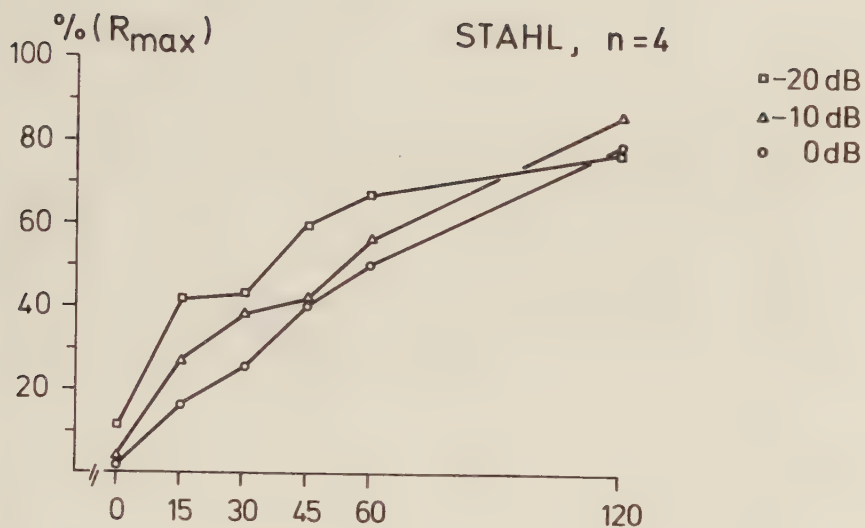


Abb. 11 : Relative Antwortraten (Erholungskurven) bei /STAHl/ und /ST/ für die drei Maskierpegel.

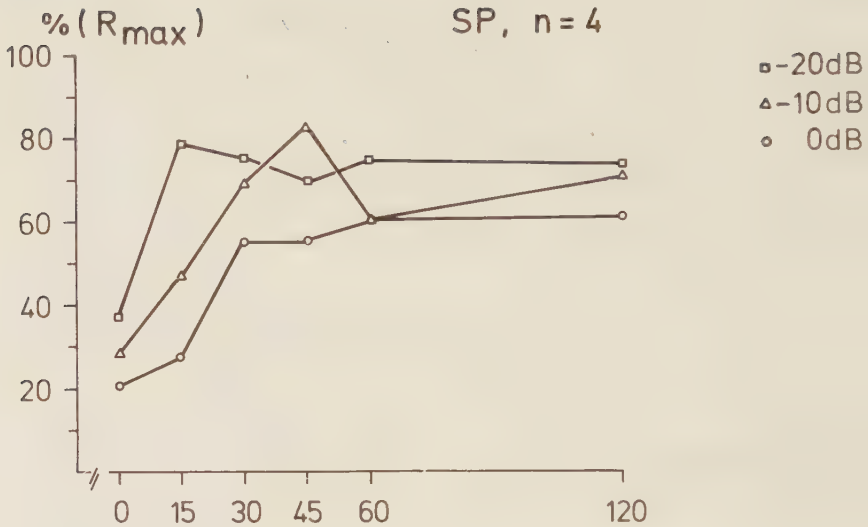


Abb. 12 : Relative Antwortraten (Erholungskurven)
bei /SP/ für die drei Maskiererpegel.

Die Erholungskurven für /STAHL/ und /ST/ sind in Abb. 11, die für /SP/ in Abb. 12 dargestellt. Bezugswert (100 %) ist jeweils die höchste bei allen Pegeln und Intervallen aufgetretene Rate des Neurons. Die Kurven zeigen die berechneten Mittelwerte.

Insgesamt zeigt sich eine verlangsamte Erholung der Antwortraten mit wachsenden Pegeln.

Für Verzögerungszeiten bis 60 ms gibt eine lineare Anpas-

Für Verzögerungszeiten bis 60 ms gibt eine lineare Anpassung die Erholungskurven bei /STAHL/ und /ST/ gut wieder (zwischen 85 % und 96 % aufgeklärte Varianz), bei Einbeziehung des 120 ms-Intervalls ist die Anpassung deutlich schlechter. Die Steigungen der Regressionsgeraden sind über die Pegel konstant, der Effekt besteht im Wesentlichen in einer Verschiebung der kritischen Pausenlängen zu höheren Werten, wenn der Maskiererpegel angehoben wird.

Diese Resultate entsprechen abgesehen vom hier linearen Verlauf der Erholungskurven bei kurzen Intervallen den andersorts berichteten Befunden (Harris & Dallos, 1979, Schreiner, 1980 a), denen zufolge höhere Maskiererpegel verlangerte Zeitkonstanten bewirken. Die dort ebenfalls berichtete Zunahme der Verzögerungszeit, bei der keine Reduktion der Antwortrate mehr auftritt, deutet sich in Abb. 11 bei den Kurven für /ST/ an. Hier liegen die mittleren relativen Raten bei den höheren Pegeln auch bei 120 ms noch deutlich unter dem Bezugswert für -20 dB.

Quantitativ von sonstigen Messungen abweichende Resultate ergeben sich bei einem Intervall von 0 ms. Die relativen Raten liegen für /STAHL/ und /ST/ mit maximal 15 % der größten Antwort unter den im ersten Versuch gemessenen, als Folge ergeben sich zwangsläufig geringere Unterschiede bei verschiedenen Maskiererpegeln.

5.2.3 Antwortverhalten einzelner Zellen

Die charakteristische oder Bestfrequenz und die Frequenzschwellenkurve konnten jeweils anschließend an die Darbietung der Sprachsignale für insgesamt 21 Zellen ermittelt werden. Zu diesem Zweck werden für Töne verschiedener Frequenz in Schritten von 3 dB die Amplituden ermittelt, bei denen die Zelle nicht mehr sichtbar antwortet. Der nächsthöhere Pegel wird als Frequenzschwelle betrachtet. Von diesen Zellen antworteten 10 ausschließlich auf /P/, 7 nur auf /T/, 4 Zellen auf beide. Abb. 13 zeigt die Häufigkeiten der Bestfrequenzen für Klassen von 1 kHz Breite.

Mit der Ausnahme eines Neurons liegen sie alle innerhalb des Spektrums der Signale. Die Bestfrequenzen für auf /T/ antwortende Zellen liegen der Tendenz nach höher als für /P/, die der auf beide antwortenden eher niedrig (0.9, 1.4, 2.0, 3.25 kHz).

Als Maß für die Frequenzselektivität der Neurone wurde der Quotient aus der Bestfrequenz und der Weite des Frequenzbereichs ermittelt, innerhalb dessen die Zellen 10 dB über ihrer Schwelle reagieren ($Q_{10\text{ dB}}$).

Es ergeben sich keine systematischen Zusammenhänge mit den Zellantworten. Die auf /P/ und /T/ gleichermaßen reagierenden Neurone sind über vergleichbare Frequenzbereiche empfindlich, wie andere entsprechender Bestfrequenz, die nur auf eine der beiden Stimulusgruppen antworten.

Weiterhin sind keine systematischen Beziehungen dieser Kennwerte zu den mit Sprachstimuli erhobenen Antwortraten festzustellen, weder bezüglich ihrer Höhe noch ihrer Abhängigkeit von den Intervallängen.

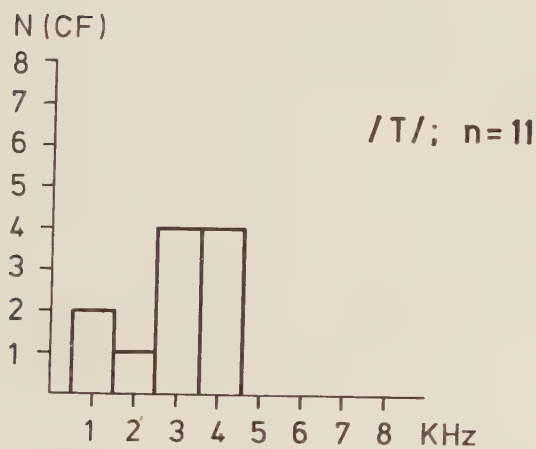
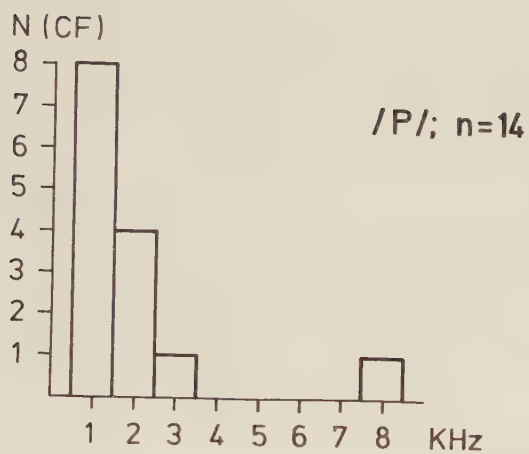


Abb. 13 : Verteilung der Bestfrequenzen

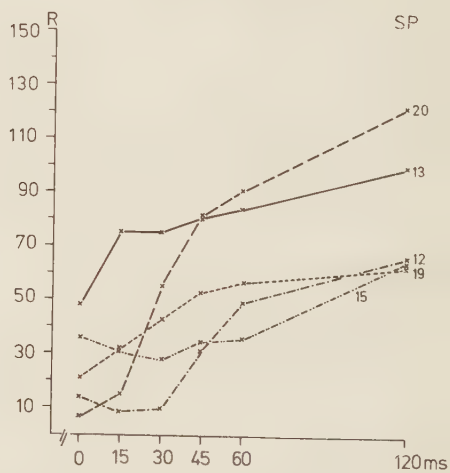
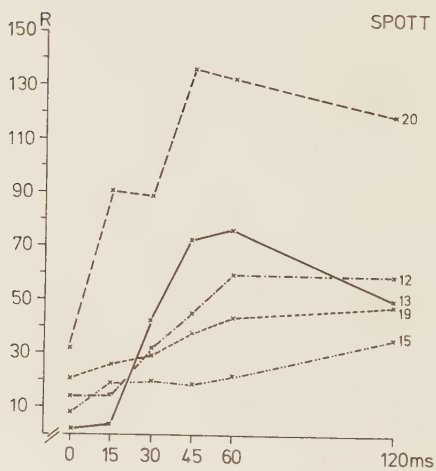
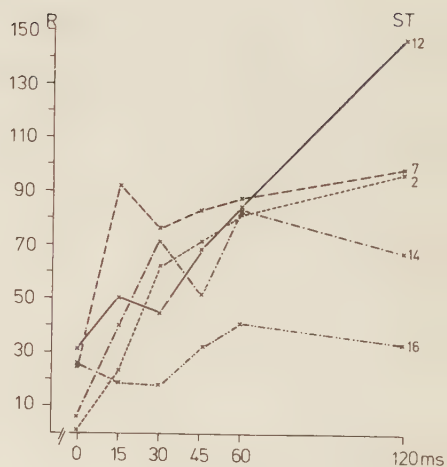
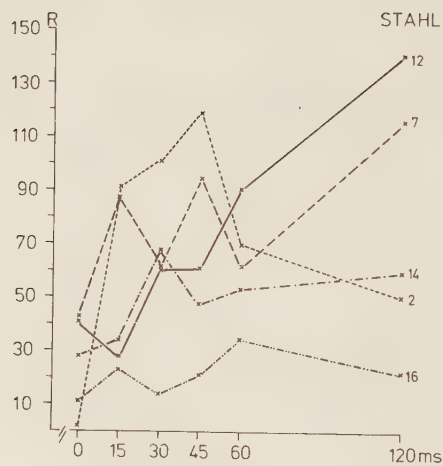


Abb. 14 : Korrigierte Antwortraten für je 5 einzelne Zellen

Die große Variabilität der Antwortkurven (vgl. Abb.14) wurde auch bei Verwendung nichtsprachlicher Reize gefunden.

Urbas & Spenner (1982) präsentierten einen Testtonimpuls bei Bestfrequenz des jeweiligen CGM-Neurons, während der Maskiertonimpuls über verschiedene Frequenzen innerhalb und außerhalb des Antwortbereichs der Zelle variiert wurde.

Die Zellantworten fallen auch hier - bei identischen Stimulusparametern relativ zur Frequenzschwellenkurve - äußerst verschiedenartig aus. Es finden sich Erholungskurven, die den mit Sprachreizen gewonnenen entsprechen, aber auch Neurone, die erst bei deutlich längeren Verzögerungszeiten als 120 ms keine Verdeckungseffekte mehr aufzeigen.

Für beide Gruppen von Stimuli zeigt sich also die von Harris & Dallos (1979) am Hörnerven festgestellte Unabhängigkeit der Erholungskurven von der Bestfrequenz auch im CGM von Katzen, bei allerdings augenscheinlich komplexerem Antwortverhalten.

Die Darstellung der Sprachreize ist von Neuron zu Neuron unterschiedlich. Sie antworten alle mit ausreichender Rate auf den Verschußlautanteil, was ja das Kriterium für ihre Auswahl war. Die Repräsentation der umgebenden Wortteile hingegen ist von sehr variabler Güte. So antwortet eine Anzahl von Zellen überhaupt nicht auf Frikative und/oder Vokale, einige übertragen die Zeitstruktur der Signale recht detailgetreu (wie das Neuron in Abb. 5).

Bei den vorhandenen Vokalantworten fällt auf, daß unabhängig von der Bestfrequenz die Zellantworten ca. alle 8 ms ein Maximum im Histogramm aufweisen, also synchron zur Grundfrequenz des Vokals (ca. 125 Hz) feuern.

Dies ist eine Entsprechung zu den Resultaten von Watanabe & Sakai (1978), die im Colliculus inferior der Katze Zellantworten auf natürliche japanische Sprache ableiteten. Ähnliches fand Hashimoto (1980) im CGM des Meerschweinchens für eine natürlich gesprochene japanische Silbe, interessanterweise nicht aber für synthetische Sprachlaute.

Bezogen auf neurophysiologische Verdeckung bestätigt sich hier, daß auf höheren Ebenen der Hörbahn zusätzlich zu Adaptationsvorgängen aktive Hemmungsprozesse anzunehmen sind, weil hier die Verringerung der Antworten auf den Testreiz nicht von einer Antwort auf das verdeckende Signal abhängt.

6. Diskussion

In beiden Versuchsreihen finden sich durchweg längere kritische Intervalle bei /SPOTT/ als bei /STAHL/.

Daß zur Wahrnehmung des bilabialen /P/ längere Verzögerungszeiten nötig sind, entspricht den Messungen anderer Autoren und den artikulatorischen Verhältnissen, da die Dauer des Verschlusses des Vokaltraktes bei der Produktion des /T/ kürzer ausfällt.

Es finden sich in diesem Zusammenhang aber auch akustische Signaleigenschaften, die mit der Annahme auditorischer Verdeckung vereinbar sind.

Abb. 15 zeigt Kurzzeitspektren (ca. 30 ms) für Ausschnitte der Frikative und der Vokalanfänge und für die Verschlusslautreize, darunter sind jeweils die zugehörigen Zeitsignale abgebildet. Die Amplitudenspektren sind dabei relativ zu der Frequenzkomponente aufgetragen, bei der beim gegebenen Signal(-ausschnitt) die maximale Energie auftritt.

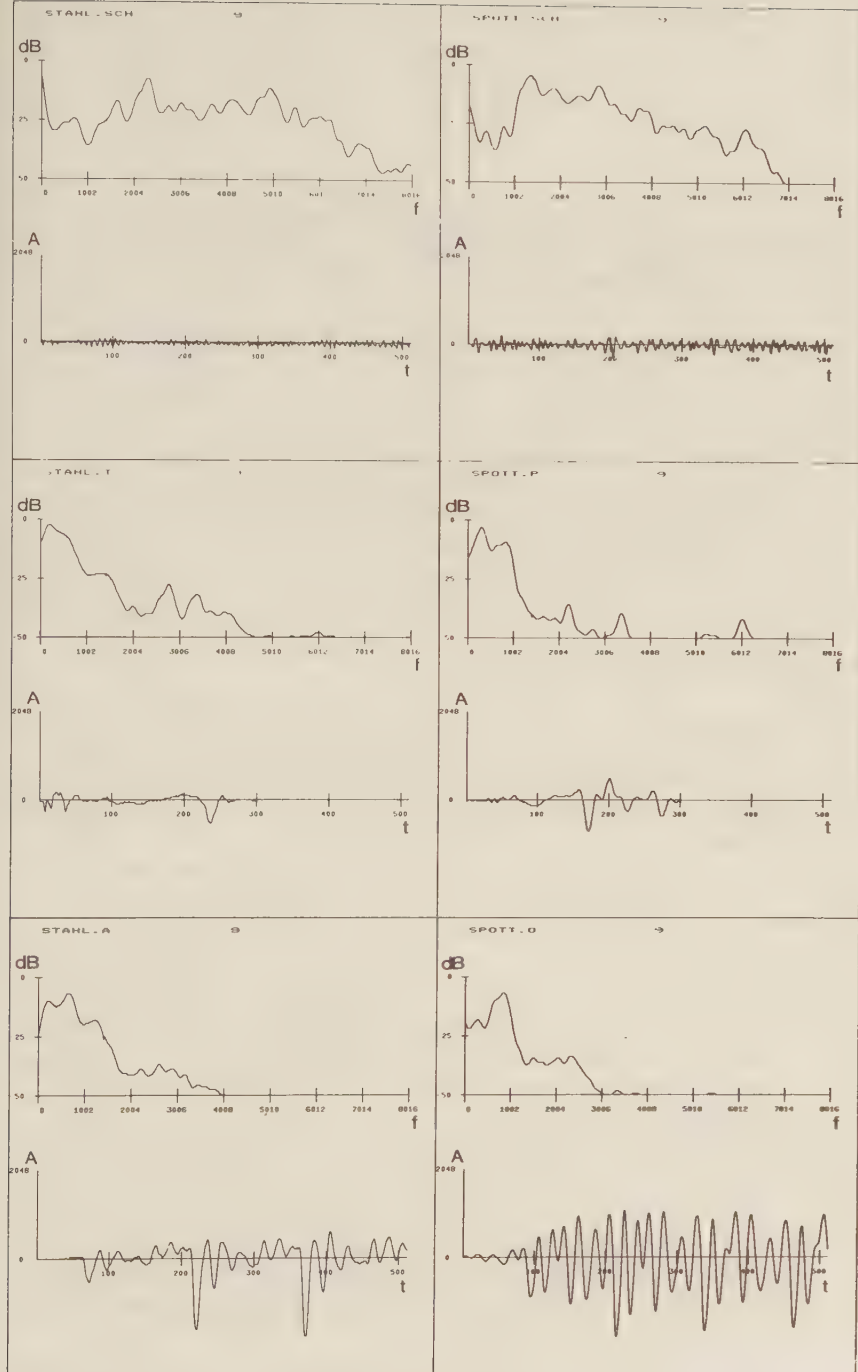


Abb. 15 : Kurzzeitspektren der (von oben nach unten) Frikative, Verschlusslaute und Vokale für /STAHL/ (links) und /SPOTT/.

Im Gegensatz zu den Amplituden der Zeitsignale sind somit die der Spektren zwischen den Sprachlautklassen nicht vergleichbar.

Die den Stimuli der ersten Versuchsreihe entnommenen Frikative (oben) decken beide einen recht breiten Bereich des Spektrums ab, mit maximaler Energie jeweils über 1 kHz. Bei den Vokalen (unten) zeigen sich die den Formanten entsprechenden spektralen Maxima, am ausgeprägtesten bei den zweiten Formanten. Die Konsonant-Vokalübergänge enthalten die niederfrequenten Komponenten der Formanttransienten, die die folgenden Vokale einleiten, und zeigen im höheren Frequenzbereich annähernd die unterschiedlichen spektralen Eigenschaften, wie sie von Stevens & Blumstein (1978) beschrieben wurden.

Die Verdeckung der den Frikativen folgenden Signalteile sollte im wesentlichen für Frequenzkomponenten über 1kHz effektiv sein, also die hochfrequenten Anteile der Verschlußlaute betreffen, und die Frequenzbereiche unter 1 kHz gering beeinflussen. In diesem Bereich haben die Formanten des /A/ und damit die Transienten bei /T/ mehr Energie als ihre Entsprechungen bei /O/ und /P/, was also weniger ausgeprägte Verdeckungseffekte und damit kürzere kritische Verzögerungszeiten erwarten läßt. Vergleichbares fanden Best et al. (1981) und auch Bailey & Summerfield (1980) mit synthetischen Stimuli : Wenn der erste Formant im Spektrum niedriger liegt, werden kürzere notwendige Pausenlängen gemessen. Auch die Energieverhältnisse in höheren Frequenzbereichen stehen einer Verdeckungsannahme nicht entgegen. Bei /STAHL/ ist die relative Energie des angenommenen Maskierers in einem Frequenzbereich, in dem der Verschlußlaut Amplitudenmaxima aufweist, im Vergleich zu den Verhältnissen bei /SPOTT/ geringer.

Eindeutig für auditorische Nachverdeckung sprechen die Ergebnisse bei den Wortstimuli mit verschiedenen Pegeln der Frikative, zumal es keinen Hinweis dafür gibt, daß in natürlicher Sprache der Vokaltrakt länger geschlossen bleibt, wenn die dem Verschußlaut vorangehenden Frikative lauter ausfallen.

Die Verschiebung zu längeren notwendigen Verzögerungszeiten mit wachsendem Pegel ist dem Verhalten von Mithörschwellen im klassischen Verdeckungsexperiment vergleichbar. So sind für einen festen Testtonpegel und variable Maskiererpegel die Mithörschwellen in gleicher Weise gegeneinander in der Zeit verschoben. Beispielsweise wird bei Zwicker (1982) bei einem Testton von 70 dB und maskierendem Rauschen von 60 und 80 dB die Mithörschwelle bei ca. 20 ms resp. ca. 40 ms Verzögerungszeit erreicht. Wenn auch Signale und Lautheitsverhältnisse mit den hier gegebenen Verhältnissen nicht übereinstimmen, so ist doch die qualitative Ähnlichkeit eindeutig.

Die neurophysiologischen Ergebnisse sind insgesamt als Verdeckungseffekte zu interpretieren. Sowohl die Wirkung verschiedener Verzögerungszeiten als auch die der Maskiererpegel zeigen qualitative und auch quantitative Parallelen zu den Befunden mit einfacheren Signalen.

Gleichzeitig findet sich eine durchgängige qualitative Übereinstimmung mit den psychoakustischen Messungen bei den Wortstimuli. Quantitative Ähnlichkeiten betreffen den Bereich relevanter Verzögerungszeiten in der ersten Versuchsreihe, wo die Zellantworten bei 60 ms Intervalllänge als erholt gelten können. Bei der Pegelvariation zeigt sich das bekannte, zwischen psychoakustischer und neurophysiologischer Verdek-

kung unterschiedliche Verhalten der Erholung, nämlich bei den Zellantworten deren Verschiebung zu längeren Pausen, als sie psychophysisch noch Effekte zeigen.

Inwieweit die aufgezeigten Verdeckungseffekte eine Unterdrückung von spektralen Komponenten des CV-Überganges bedeuten, also der Verschlußlaut wegen mangelhafter Codierung nicht übertragen wird, ist nicht zu klären, weil die Analyse der Einzelzellantworten keine Aussage darüber zulaßt, welche Signalkomponenten dargestellt werden.

Alternativ wäre eine Erklärung als Unterdrückung der zeitlichen Markierung von Signalteilen möglich, etwa im Sinne der Resultate von Schreiner (1980 b, s.o.) mit alternierenden Signalen.

Schwieriger ist die Interpretation der Ergebnisse für die Verschlußlautreize, die wie erwähnt auch dann gehört werden, wenn sich keine Pause zwischen Frikativen und Silbenanfangen befindet.

Andererseits zeigen die abgeleiteten Neurone deutliche Nachverdeckungseffekte, die bis zu 60 ms intervallange quantitativ den bei den Wortreizen gefundenen sehr gut entsprechen.

Akustisch enthalten die Verschlußlautreize schnelle spektrale Änderungen mit wenig Energie und relativ hohen Frequenzkomponenten, die bei der Öffnung des Vokaltraktes entstehen, und zumindest Anteile der spektral dem Vokal verwandten Transienten (vgl. Abb. 15). Letztere sind deutlich energiereicher und befinden in zeitlich größerem Abstand zu den Frikativen. So sollten die Frikative weniger effektiv die vokalischen Signalkomponenten maskieren, als die akustisch 'näheren' erstgenannten Anteile.

So könnte erklärt werden, daß man den Silbenbeginn als Geräusch hört, ihm aber keine Sprachqualität zuschreibt.

Bei den Wortstimuli werden diese von den Frikativen nicht verdeckten Anteile dann möglicherweise durch den folgenden Vokal unterdrückt, da sie diesem gegenüber kurz und energiearm sind. So fanden Resnick et al. (1979) Vor- und Nachverdeckung von 25 ms langen Rauschimpulsen durch synthetische Vokalannäherungen, deren Ausmaß von der spektralen Ähnlichkeit von Test- und Maskierreiz abhängt. Ihrer Meinung nach jedoch spielt insbesondere die Vorverdeckung wegen insgesamt geringer Beträge nur eine untergeordnete Rolle bei der Wahrnehmung von Verschußlauten.

Die neurophysiologischen Ergebnisse liefern einen Hinweis auf Vorverdeckung, sofern sie für /ST/ und /SP/ längere notwendige Verzögerungszeiten ergeben, da die Antwortraten hier bis 120 ms im Gegensatz zu denen bei Wortreizen noch zunehmen. Dies muß jedoch eine Vermutung bleiben, da eine Verminderung der Antwort auf den Verschußlautanteil durch diejenige auf den Vokalteil nur bei vereinzelten Zellen direkt gemessen wurde.

Aus dieser Sicht ist dann der höherfrequente, rauschhafte Anteil am Beginn von CV-Silben für die Wahrnehmung der Verschußlaute von wesentlicher Bedeutung.

Dies unterstützt der Befund von Sharf & Hemeyer (1972), wonach bei natürlichen Stimuli Formanttransienten allein zur Verschußlauterkennung nicht ausreichen. Auch ein Diskussionsbeitrag von Pastore (1981) zu den Arbeiten von Stevens & Blumstein (1978, s.o.) ist in diesem Zusammenhang von Interesse : Abweichend von den neurophysiologischen Resultaten von Möller (1974, s.o.) davon ausgehend, daß schnell

veränderliche Signale vom menschlichen Gehör weniger exakt spektral aufgelöst und dargestellt werden als eher kontinuierliche, vermutet er, daß sie in Sprachsignalen evtl. die Funktion haben, die ihnen folgenden Transienten als solche zu 'markieren', also darauf hinzuweisen, daß ein Verschußlaut vorliegt. (Hierzu würde dann die zweite Hypothese zur neurophysiologischen Verdeckung passen).

Anders als die qualitativen, die als gesichert gelten dürfen, sind quantitative Übereinstimmungen entweder nicht deutlich genug ausgeprägt oder aber mit Zurückhaltung zu interpretieren.

Die Ermittlung psychoakustischer Verdeckungsmaße scheitert zum einen an der Komplexität der Sprachsignale. Diese sind zwar, z.B. durch Filterung, in einfache Komponenten zerlegbar, für die dann klassische Angaben in Form von Mithörschwellen möglich wären. Es ist aber a priori schwer entscheidbar, welche Komponenten die relevanten für einen spezifischen Sprachlaut, insbesondere einen Konsonanten darstellen. Hier ergibt sich die Notwendigkeit weiterer Erforschung der akustischen Merkmale von Sprachsignalen, auch mit dem Ziel, Signaltypen zu erhalten, die graduell unterschiedlich komplex und damit für klassische psychoakustische Untersuchungen geeignet sind.

Zum zweiten bedeutet die Wahrnehmung von Sprachlauten ihre Zuordnung zu phonetischen Klassen ('categorical perception', s.o.), die akustischen Verhältnisse hingegen zeigen kontinuierliche Veränderungen. Dies führt zu unterschiedlichen Maßen bei der Untersuchung der Wahrnehmung von Sprachlauten gegenüber denen mit einfachen (kontinuierlicher Variation zugänglichen) Signalen in der Psychoakustik, die nicht ohne weiteres ineinander überführt werden können. Es kann somit nicht

entschieden werden, ob die gefundenen Verdeckungseffekte bei den Sprachreizen in ihrem Ausmaß so bedeutsam sind, daß sie die Unterdrückung von Verschluslauten vollständig erklären.

Über die bestehenden Anstrengungen hinaus (z.B. Macmillan et al., 1977, Pastore, 1981) sollten also Wege gesucht werden, die Meßmethoden und damit die Maße vergleichbar zu machen.

An den neurophysiologischen Ergebnissen fällt auf, daß sich die Antwortraten in Verdeckungsexperimenten mit komplexen Sprachreizen und einfachen Signalen quantitativ durchaus ähnlich verhalten.

Die Bedenken gegenüber quantitativen Angaben betreffen daher im Wesentlichen ihre Vergleichbarkeit mit psychoakustischen Werten. Wie bereits erwähnt, beruhen diesbezügliche Einschränkungen u.a. darauf, daß Antworten kleiner Subsysteme (Zellen) mit einer Antwort des Gesamtsystems verglichen werden.

Bestehen auch keine Zweifel, daß Zellantworten mit Wahrnehmungsleistungen insofern zusammenhängen, als höhere Antwortraten eine höhere Wahrscheinlichkeit des Zustandekommens einer bestimmten Wahrnehmung zur Folge haben, so sind doch keine exakten vergleichenden Aussagen möglich.

Auch hierzu bietet sich an, psychoakustische und neurophysiologische Experimentier- und Meßverfahren einander genügend anzugleichen, ein Ziel, das aus technischen Gründen nur schwer zu erreichen sein wird. Eine andere, experimentell aufwendige Möglichkeit ist die Kopplung von neurophysiologischen Ableitungen mit Verhaltensexperimenten am selben Versuchstier oder solchen derselben Species. Diese könnten Hinweise geben, wie Zellaktivität und Wahrnehmungsreaktionen bzw. deren Änderungen zusammenhängen. Bei Ableitung am wachen

Tier würden dabei auch die evtl. verzerrenden Einflüsse der Anästhesie entfallen.

Trotz dieser und anderer Einwände (z.B. zur Variabilität innerhalb und zwischen verschiedenen Species) sind jedoch die bisher gefundenen diversen Übereinstimmungen (vgl. z.B. Evans, 1977) von psychoakustischen und neurophysiologischen Ergebnissen zum Gehör von Säugetieren recht vielversprechend, die hier zu Verdeckungsprozessen vorliegenden eingeschlossen.

Wenn also auch einiges offen bleiben muß, so ist doch die Durchgängigkeit der qualitativen Übereinstimmungen zwischen psychoakustischen Messungen bei den Wortreizen mit neurophysiologischer Verdeckung ein starkes Argument für eine Beteiligung von auditorischer Maskierung an der Entstehung des untersuchten Phänomens.

7. Zusammenfassung

Einige Ansätze zur Wahrnehmung von Sprachlauten postulieren sprachspezifische phonetische Prozesse, während andere davon ausgehen, daß die Eigenschaften des auditorischen Systems ausreichende Erklärungen liefern können.

Ein Phänomen, das als Beleg für sprachspezifische, auf die Sprachproduktion bezogene Prozesse (Motor Theory of Speech Perception) herangezogen wird, ist die sogenannte 'Verschlußlautunterdrückung': Entfernt man das dem Verschluß des Vokaltraktes entsprechende Intervall im Sprachsignal, wird der - akustisch vorhandene - Verschlußlaut nicht gehört, aus z.B. /STAHL/ wird /SCHAL/. Es wird nun angenommen, daß das Fehlen der artikulatorischen Referenz (des Verschlusses) zu diesem Phänomen führt.

Andererseits entspricht die Anordnung der verschiedenen Wortteile dem psychoakustischen Paradigma der Nachverdeckung: Ein (lauter) Störschall führt zu verminderter Hörbarkeit eines folgenden Schalles. Da diese auditorische Hypothese wegen der Komplexität der Sprachsignale nicht mit klassischen psychoakustischen Methoden prüfbar ist, und andererseits Tierversuche die Wirkung sprachspezifischer Wahrnehmungsvorgänge ausschließen, werden psychophysische Resultate mit denen aus neurophysiologischen Versuchen verglichen.

Die deutschen, natürlich gesprochenen Wörter /STAHL/ und /SPOTT/ werden digital mit verschiedenen langen Intervallen zwischen dem frikativen /SCH/ und dem jeweiligen Verschlußlaut (/T/, /P/) versehen, außerdem wird der Pegel der /SCH/ variiert. Im psychoakustischen Experiment wird untersucht, bei welcher Intervalllänge der Verschlußlaut gehört wird.

Im neurophysiologischen Versuch werden im Corpus geniculatum mediale der Katze extrazellulär Antworten von Einzelzellen abgeleitet und die Antwortraten für die verschiedenen Intervalle ermittelt.

Es zeigt sich, daß

1. die akustischen Eigenschaften der Wortteile und die psychophysischen Meßergebnisse einer Nachverdeckungsannahme nicht widersprechen,

2., daß psychophysische Messungen mit verschiedenen Pegeln der frikativen /SCH/ diese Annahme deutlich unterstützen, daß

3. die neurophysiologischen Ergebnisse qualitativ und quantitativ denen zur neurophysiologischen Nachverdeckung mit einfachen Signalen entsprechen, und daß

4. mit Sprachreizen gewonnene psychoakustische und neurophysiologische Befunde qualitativ übereinstimmen, wie es auch für einfache Signale festgestellt wurde.

Aus diesen Entsprechungen wird geschlossen, daß auditorische Verdeckungsprozesse bei der Wahrnehmung von Verschlußlauten eine Rolle spielen, der Bezug auf sprachspezifische Vorgänge der Produktion hier also nicht notwendig erscheint.

8. Literaturverzeichnis

Aitkin, L.M. & Dunlop, C.W. (1969) : Inhibition in the Medial Geniculate Body of the Cat. *Exp. Brain Res.* 7, 68 - 83.

Bailey, P.J. & Summerfield, Q. (1980) : Information in Speech : Observations on the Perception of /S/-Stop Clusters. *J. Exp. Psych. Hum. Perc. Perf.* 6, 536 - 563.

Best, C.T., Morrongiello, B., Robson, R. (1981) : Perceptual Equivalence of Acoustic Cues in Speech and Nonspeech Perception. *Perc. Psychophys.* 29, 191 - 211.

Blumstein, S.E. & Stevens, K.N. (1979) : Acoustic Invariance in Speech Production : Evidence from Measurements of the Spectral Characteristics of Stop-Consonants. *J. Acoust. Soc. Am.* 66, 1001 - 1017.

Blumstein, S.E. & Stevens, K.N. (1980) : Perceptual Invariance and Onset Spectra for Stop-Consonants in Different Vowel Environments. *J. Acoust. Soc. Am.* 67, 648 - 662.

Creutzfeldt, O.D., Hellweg, F.C., Schreiner, Chr. (1980) : Thalamocortical Transformation of Responses to Complex Auditory Stimuli. *Exp. Brain Res.* 39, 87 - 104.

Creutzfeldt, O.D. & Probst, W. (1974) : Neurophysiologische Technik. In : Gauer, O.H., Kramer, K., Jung, R. (Hrsg.) : *Physiologie des Menschen*. Bd. 10 : Allgemeine Neurophysiologie, 2. Aufl.. Urban & Schwarzenberg, München.

David, E., Keidel, W.D., Kallert, S., Bechterova, N.P., Bundsen, P.V. (1977) : Decoding Processes in the Auditory System and Human Speech Analysis. In : Evans, E.F. & Wilson, J.P. (Hrsg.) : Psychophysics and Physiology of Hearing. Academic Press, London.

Delgutte, B. (1980) : Representation of Speech-Like Sounds in the Discharge Patterns of Auditory-Nerve Fibers. J. Acoust. Soc. Am. 68, 843 - 857.

Diehl, R.L. (1981) : Feature Detectors for Speech : A Critical Reappraisal. Psych. Bull. 89, 1 - 18.

Dorman, M.F., Raphael, L.J., Liberman, A.M. (1979) : Some Experiments on the Sound of Silence in Phonetic Perception. J. Acoust. Soc. Am. 65, 1518 - 1532.

Evans, E.F. (1975) : Cochlear Nerve and Cochlear Nucleus. In : Keidel, W.D. & Neff, W.D. (Hrsg.) : Handbook of Sensory Physiology, Vol. V/2, Auditory System. Springer-Verlag, Berlin.

Evans, E.F. (1977) : Some Interactions between Physiology and Psychophysics in Acoustics. In : Proc. 9 th Int. Congress on Acoustics; Volume of Invited Review Lectures, 55 - 65, Spanish Acoustical Society, Madrid.

Evans, E.F. (1981 a) : The Dynamic Range Problem : Place and Time Coding at the Level of Cochlear Nerve and Nucleus. In : Syka, J. & Aitkin, L. (Hrsg.) : Neuronal Mechanisms of Hearing. Plenum Press, New York.

Evans, E.F. (1981 b) : Neural Encoding of Speech Signals at Peripheral and Central Levels of the Auditory System. In : Myers, T., Laver, J., Anderson, J. (Hrsg.) : The Cognitive Representation of Speech. North Holland Publ. Co., Elsevier.

Harris, D.M. & Dallos, P. (1979) : Forward Masking of Auditory Nerve Fiber Responses. J. Neurophys., 42, 1083 - 1107.

Hashimoto, T. (1980) : Information Processing of Speech Sounds in the Medial Geniculate and the Inferior Colliculus. Proc. Japan. Acad. 56, Ser. B, 294 - 299.

Karnickaya, E.G., Mushnikov, V.N., Slepokurova, N.A., Zhukov, S.Ja. (1975) : Auditory Processing of Steady State Vowels. In : Fant, G. & Tatham, M.A.A. (Hrsg.) : Auditory Analysis and Perception of Speech. Academic Press, London.

Klatt, D.H. (1979) : Speech Perception : a Model of Acoustic-Phonetic Analysis and Lexical Access. J. Phonetics 7, 279 - 312.

Kuhl, P.K. (1979) : Models and Mechanisms in Speech Perception. Species Comparisons Provide Further Contributions. Brain Behav. Evol. 16, 374 - 408.

Kuhl, P.K. (1981) : Discrimination of Speech by Nonhuman Animals : Basic Auditory Sensitivities Conducive to the Perception of Speech-Sound Categories. J. Acoust. Soc. Am. 70(2), 340 - 349.

Ladefoged, P. (1975) : A Course in Phonetics. Harcourt Brace Jovanovich, New York.

Lewien, T. (1983) : Filterung von Spracheinhüllenden zur Verständlichkeitsverbesserung bei Innenohrschwerhörigkeit. Dissertation, Göttingen.

Lenneberg, E.H. (1967) : Biological Foundations of Language. Wiley, New York.

Lieberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M. (1967) : Perception of the Speech Code. Psych. Rev. 74, 431 - 461.

Lieberman, A.M. & Pisoni, D.B. (1976) : Evidence for a Special Speech Perceiving Subsystem in the Human. In : Bullock, Th.E. (Hrsg.) : Dahlem Workshop on Recognition of Complex Acoustic Signals. Dahlem Konferenzen, Berlin.

Lieberman, A.M. & Studdert-Kennedy, M. (1978) : Phonetic Perception. In : Held, R., Leibowitz, H.W., Teuber, H.-L. (Hrsg.) : Handbook of Sensory Physiology, Vol. VIII, Perception. Springer-Verlag, Berlin.

Lienert, G.A. (1973) : Verteilungsfreie Methoden in der Biostatistik, Bd. 1, 2.Aufl., Hain-Verlag, Meisenheim.

Macmillan, N.A., Kaplan, H.L., Creelman, C.D. (1977) : The Psychophysics of Categorical Perception. Psych. Rev. 84, 452 - 471.

Marcus, S.M. (1978) : Distinguishing "slit" and "split" - an Invariant Timing Cue in Speech Perception. Perception & Psychophysics 23, 58 - 60.

Massaro, D.W. (1972) : Preperceptual Images, Processing Time, and Perceptual Units in Auditory Perception. Psych. Rev. 79, 124 - 145.

Massaro, D.W. (1975) : Acoustic Features in Speech Perception. In : Massaro, D.W. (Hrsg.) : Understanding Language. An Information Processing Analysis of Speech Perception, Reading and Psycholinguistics. Academic Press, New York.

Møller, A.R. (1974) : Coding of Sounds with Rapidly Varying Spectrum in the Cochlear Nucleus. J. Acoust. soc. Am. 55, 631 - 640.

Paap, K.R. (1975) : Theories of Speech Perception. In : Massaro, D.W. (Hrsg.) : Understanding Language. An Information Processing Analysis of Speech Perception, Reading and Psycholinguistics. Academic Press, New York.

Pastore, R.E. (1981) : Possible Psychoacoustic Factors in Speech Perception. In : Eimas, P.D. & Miller, J.L. (Hrsg.) : Perspectives on the Study of Speech. Erlbaum, Hillsdale, N.J..

Pisoni, D.B. (1979) : On the Perception of Speech Sounds as Biologically Significant Signals. Brain Behav. Evol. 16, 330 - 350.

Resnick, S.B., Weiss, M.S., Heinz, J.M. (1979) : Masking of Filtered Noise Bursts by Synthetic Vowels. J. Acoust. Soc. Am. 66, 674 - 677.

Sachs, L. (1972) : Statistische Auswertungsmethoden, 3. Aufl. Springer - Verlag, Berlin.

Sachs, M.B. & Young, E.D. (1979) : Encoding of Steady-State Vowels in the Auditory Nerve : Representation in Terms of Discharge Rate. J. Acoust. Soc. Am. 66, 470 - 479.

Schreiner, C. (1979) : Temporal Suppression and Speech Processing. In : Creutzfeldt, O., Scheich, H., Schreiner, C. (Hrsg.) : Hearing Mechanisms and Speech. Exp. Brain Res., Suppl.II, Springer-Verlag, Berlin.

Schreiner, C. (1980 a) : Nachverdeckung in zentralen Stationen der Hörbahn. In : Fortschritte der Akustik - DAGA '80. VDE-Verlag, Berlin.

Schreiner, C. (1980 b) : Encoding of Alternating Acoustical Signals in the Medial Geniculate Body of Guinea Pigs. Hear. Res. 3, 265 - 278.

Sharf, D.J. & Hemeyer, T. (1972) : Identification of Place of Consonant Articulation from Vowel Formant Transitions. J. Acoust. Soc. Am. 51, 652 - 658.

Sinex, D.G. & Geisler, C.D. (1983) : Responses of Auditory-Nerve Fibers to Consonant-Vowel Syllables. J. Acoust. Soc. Am. 73, 602 - 615.

Sixtl, F. (1967) : Meßmethoden der Psychologie. Beltz, Weinheim.

Snowdon, C.T. (1979) : Response of Nonhuman Animals to Speech and to Species-Specific Sounds. Brain Behav. Evol. 16, 409 - 429.

Spenner, H. & Schreiner, C. (1980) : Stop-Konsonanten und Zeitdauer des vorangehenden Intervalls im Sprachsignal : Psychophysische und Neurophysiologische Messungen. In : Fortschritte der Akustik. DAGA '80. VDE-Verlag, Berlin.

Stevens, K.N. (1980) : Acoustic Correlates of Some Phonetic Categories. J. Acoust. Soc. Am. 68, 836 - 842.

Stevens, K.N. & Blumstein, S.E. (1979) : Acoustic Invariance in Speech Production : Evidence from Measurements of the Spectral Characteristics of Stop-Consonants. J. Acoust. Soc. Am. 66, 1001 - 1017.

Tekieli, M.E. & Cullinan, W.L. (1979) : The Perception of Temporally Segmented Vowels and Consonant-Vowel Syllables. J. Speech Hear. Res. 22, 88 - 102.

Urbas, J.V. & Spenner, H. (1982) : Cat Medial Geniculate Body (MGB) Neuron Responses to Two-Tone Forward Masking Paradigms Using Human Speech or Tone Bursts. In : Fortschritte der Akustik - FASE/DAGA '82. VDE-Verlag, Berlin.

Watanabe, T. & Sakai, H. (1978) : Responses of the Cat's Collicular Neuron to Human Speech. J. Acoust. Soc. Am. 64, 333 - 337.

Young, E.D. & Sachs, M.B. (1979) : Representation of Steady-State Vowels in the Temporal Aspects of the Discharge Patterns of Populations of Auditory-Nerve Fibers. J. Acoust. Soc. Am. 66, 1381 - 1403.

Zwicker, E. (1982) : Psychoakustik. Springer-Verlag, Berlin.

Zwicker, E. & Feldtkeller, R. (1967) : Das Ohr als Nachrichtenempfänger. 2.Aufl. Hirzel-Verlag, Stuttgart.

Danksagungen

Herrn Prof. Dr. O.D. Creutzfeldt danke ich für die Möglichkeit, diese Arbeit in der von ihm geleiteten Abteilung Neurobiologie des MPI für biophysikalische Chemie erstellen zu können, für seine Unterstützung und Anregungen.

Prof. Dr. M. Stadler sei bedankt für die Übernahme eines Gutachtens als Fachvertreter, die Vertretung der Arbeit an der Universität Bremen, sowie für wertvolle Hinweise.

Ganz besonders verpflichtet bin ich J.V. Urbas, der die neurophysiologischen Messungen mit mir zusammen durchführte und ein stets fruchtbarer Diskussionspartner war.

C. Schreiner danke ich für seine Mitarbeit an den Voruntersuchungen und für viele hilfreiche Gespräche.

Probleme beim Erstellen der Rechnerprogramme wurden mit Hilfe von T. Lewien gelöst, auch danke ich ihm einige wichtige Unterprogramme.

Schließlich sei den Mitarbeitern der Abteilung Neurobiologie gedankt, die in verschiedenster Weise zum Gelingen dieser Arbeit beigetragen haben.

